

Binding efficiencies of carbohydrate ligands with different genotypes of cholera toxin B: molecular modeling, dynamics and docking simulation studies

Mobashar Hussain Urf Turabe Fazil · Sunil Kumar · Rohit Farmer · Haushila Prasad Pandey · Durg Vijai Singh

Received: 29 September 2010 / Accepted: 27 December 2010 / Published online: 16 March 2011
© Springer-Verlag 2011

Abstract *Vibrio cholerae* produces cholera toxin (CT) that consists of two subunits, A and B, and is encoded by a filamentous phage CTX Φ . The A subunit carries enzymatic activity that ribosylates ADP, whereas the B subunit binds to monosialoganglioside (GM1) receptor in epithelial cells. Molecular analysis of toxigenic *V. cholerae* strains indicated the presence of multiple *ctxB* genotypes. In this study, we employed a comparative modeling approach to define the structural features of all known variants of *ctxB* found in O139 serogroup *V. cholerae*. Modeling, molecular dynamics and docking simulations studies suggested subtle variations in the binding ability of *ctxB* variants to carbohydrate ligands of GM1 (galactose, sialic acid and *N*-acetyl galactosamine). These findings throw light on the molecular efficiencies of pathogenic isolates of *V. cholerae* harboring natural variants of *ctxB* in causing the disease, thus

suggesting the need to consider *ctxB* variations when designing vaccines against cholera.

Keywords Cholera toxin B · GM1 receptor · Galactose · Molecular dynamics · Protein modeling · Docking simulations

Introduction

Vibrio cholerae is the causative agent of the deadly diarrheal disease cholera. Of the 209 serogroups, only the O1 (which is further classified into the two biotypes El Tor and Classical) and O139 serogroups of *V. cholerae* have the potential to cause epidemic and pandemic cholera [1]. The colonization, subsequent production and secretion of cholera toxin (CT) by *V. cholerae* in the intestinal cells lead to diarrhea, which can be lethal if left untreated. The CT-encoding genes, *ctxAB*, are found within a filamentous bacteriophage CTX Φ [2]. Based on the variations in the repressor protein, *rstR*, CTX Φ has been classified into three major types: CTX^{Class} Φ (classical type), CTX^{ET} Φ (El Tor type), and CTX^{Calc} Φ (Calcutta type) [1]. In *V. cholerae* O1 El Tor and O139, multiple CTX Φ prophages are site-specifically inserted near the terminus of the large chromosome, whereas classical prophages insert at the termini of both chromosomes [3, 4]. The characterization of *ctxAB* mutants presented evidence that the toxin is advantageous to the growth of the bacterium in intestinal milieu [5]. It was demonstrated that the *ctx* mutants colonized less efficiently in rabbit intestines than the parental strain. CT enhances intestinal colonization either by compensating for nutritional deficiency or by suppressing the bactericidal activity produced by epithelial cells [6].

M. H. U. T. Fazil · D. V. Singh (✉)
Infectious Disease Biology, Institute of Life Sciences,
Nalco Square,
Bhubaneswar 751023, India
e-mail: durg_singh@yahoo.co.in

S. Kumar
Bioinformatics Centre, Institute of Life Sciences,
Nalco Square,
Bhubaneswar 751023, India

R. Farmer
Department of Computational Biology and Bioinformatics, Jacob
School of Biotechnology and Bioengineering, Sam Higginbottom
Institute of Agriculture, Technology and Sciences,
Allahabad 211007, India

H. P. Pandey
Department of Biochemistry, Faculty of Science,
Banaras Hindu University,
Varanasi 221005, India

Cholera toxin is a hexameric A-B5 type toxin. The toxicity is attributed to the enzymatic activity of the A subunit, which catalyzes ADP ribosylation of the α subunit of GTP binding protein, G_s , leading to activation of adenylate cyclase and a concomitant elevation in cAMP levels, in turn causing the hypersecretion of Cl^- ions and water, and ultimately profuse diarrhea. Cholera toxin A subunit (CTA) consists of two polypeptide chains, CTA1 and CTA2. CTA1 confers CT-mediated toxigenicity, whereas CTA2 acts as a linker between CTA1 and cholera toxin B subunit (CTB). The five B subunits of the toxin bind principally to monosialoganglioside (GM1) receptors that are present on the surfaces of mammalian cells. The B subunit is considered a molecular recognition unit and delivery vehicle for the A subunit [7, 8].

Although the action of CT is conserved among classical and El Tor strains, the CTB sequence differs among the two biotypes, which serves as the basis for *ctxB* genotyping. Based on nonrandom base variations, three types of *ctxB* genes have been described, due to changes in the deduced amino acid sequence positions at 39, 46 and 68 [9]. Genotype 1 is found in classical biotype strains worldwide and in US Gulf Coast strains, genotype 2 in El Tor biotype strains from Australia, and genotype 3 in El Tor biotype strains from the seventh pandemic and the Latin American epidemic. New *ctxB* variants showing additional polymorphism at amino acid positions 28 and 34 have recently been described in *V. cholerae* O139 strains [10].

Cholera toxin B subunit is used to develop the oral cholera vaccine. When administered orally, CTB induces immunogenicity at mucosal surfaces [11, 12]. This event is believed to be the result of CTB binding to eukaryotic cell surfaces via GM1, which elicits a mucosal immune response against the pathogen and an enhanced immune response when coupled chemically with other antigens. Previous studies have indicated that amino acid sequence diversity at positions not involved in receptor binding can lead to epitope variations of CTB, and therefore must be considered during the development of synthetic and natural vaccines against cholera [13]. Mutation analysis of CTB revealed variations in immunoreactivity, hemolysis and GM1 binding ability for the toxin subunit [14, 15]. However, the role of natural variations in CTB and their advantages, if any, to this organism in terms of survival or pathogenesis have barely been explored. In this study, we carried out molecular modeling, dynamics simulations and docking studies of all natural CTB variants. Monosaccharide libraries have previously been used to explore binding sites in order to develop simple and easily synthesizable molecules against *E. coli* heat labile toxin, a close homolog of CT [16]. In this regard, monosaccharide components of GM1 receptor, namely galactose, sialic acid, N-acetyl galactosamine and glucose, were employed individually

for docking simulations. The information generated should prove useful for establishing the enterotoxic potentials of the CTB variants, if any, and for determining the relationships between variants of CTB and disease severity, aside from defining the requisite binding sites of drug-like molecules.

Materials and methods

Bacterial strains

Seven representative strains, of which five were *V. cholerae* O139 strains isolated from diarrheal patients from Kerala, southern India, were included in the study (Table 1). Reference strains of *V. cholerae* O395 (O1 classical Ogawa), N16961 (O1 Biotype El Tor) provided by Dr. R. K. Bhadra of the National Institute of Cholera and Enteric Diseases (NICED), Kolkata, were used as control. All strains were maintained in tryptic soy broth (Difco) supplemented with 30% glycerol at -80°C or in Luria Bertani (LB) agar stab culture at room temperature.

DNA isolation and sequencing

Chromosomal DNA from *V. cholerae* strains grown overnight was isolated following the method of Ausubel et al. [17]. To generate a sequencing template from strains carrying either the El Tor or the Calcutta CTX Φ , the DNA sequence between the *rstR* gene and the 3' end repeat of *attRS* was amplified. PCR products obtained from these reactions were then used as templates to get a *ctxB* gene product via the method described earlier [18]. PCR products were purified using QIAquick gel extraction kit (Qiagen) and sequences for both the DNA strands were determined using a CEQ 8000 automated DNA sequencer (Beckman Coulter, Brea, CA, USA).

Modeling of CTB variants

The experimentally generated crystal or NMR structures for all known variants in *V. cholerae* are not yet available. Therefore, we built molecular models for all the reported variants of the *ctxB* gene (Table 1) by means of comparative modeling. The crystal structure of the cholera toxin B pentamer complexed with GM1 pentasaccharide (PDB code: 2CHB) was selected as template for all modeling procedures. Sequence alignment was achieved using the Clustal W [19] (<http://www.ebi.ac.uk/clustalw>) software. Since the first 21 amino acids belonging to the N-terminal region of the target protein did not have an equivalent region in the crystal structure, modeling was carried out using the academic version of MODELER9v6

Table 1 Amino acid variations among the *ctxB* genotypes present in different strains of *V. cholerae* O1 and O139

Strains	Phage type	<i>ctxB</i> genotype	Amino acid position variation					References
			28	34	39	46	68	
Classical, 569B		Type 1	D	H	H	F	T	[10]
El Tor, Australia		Type 2	D	H	H	L	T	
El Tor, N16961		Type 3	D	H	Y	F	I	
<i>V. cholerae</i> O139, (2005)		Type 6	D	P	Y	F	T	
<i>ctxB</i> selectively amplified from CTX Φ prophages of <i>V. cholerae</i> O139								
<i>V. cholerae</i> O139 AL49	El Tor	Type 4	D	H	Y	F	T	This study
<i>V. cholerae</i> O139 AL40	El Tor	Type 5	A	H	H	F	T	
<i>V. cholerae</i> O139, AL72	Calcutta	Type 1	D	H	H	F	T	
<i>V. cholerae</i> O139, TV169	Calcutta	Type 4	D	H	Y	F	T	
<i>V. cholerae</i> O139, TV258	Calcutta	Type 5	A	H	H	F	T	

(<http://salilab.org/modeler>) [20] from amino acid positions 22 to 124 of the CTB protein. A total of 20 models were generated by MODELLER, among which the one with the best PROCHECK [21] G-score and VERIFY3D [22] profile was subjected to MD simulations and energy minimization. A nonbinding cut off of 14 Å, the CHARMM force field [23] and CHARM-all-atom charges were employed, and a steepest descent algorithm was used to remove close van der Waals contacts. This was followed by conjugate gradient minimization until the energy exhibited stability during sequential repetition. All hydrogen atoms were included during the calculation. The energy minimization started with the side chain first, and then all of the atoms of the protein were relaxed during optimization. All calculations were performed by using the ACCELRYDS modeling 2.0 software suite (Accelrys Inc., San Diego, CA, USA). VERIFY3D (a structure evaluation server) was used to check the residue profiles of the three-dimensional models obtained. In order to assess the stereochemical qualities of the three-dimensional models, PROCHECK analysis was performed. Quality evaluation of the model for the environment profile was also predicted using the ERRAT (structure evaluation server) software [24].

Molecular dynamics simulations

All atom MD simulations of CTB protein in explicit water were carried out using the GROMACS 4.0.6 software and the GROMOS96 [25] force field for a time scale of 1 ns. Three-dimensional periodic boundary conditions were imposed, enclosing the molecule in a truncated octahedron (0.5 nm thick) solvated with the SPC water model [26] provided in the GROMACS package. The system was neutralized with a single Cl⁻ counterion. The 3D molecule was locally minimized in order to provide a first minimization of the rough geometry derived from homology modeling using 1000 steps of steepest descent energy minimization.

The electrostatic terms were described using the particle mesh Ewald algorithm. The LINCS [27] algorithm was

used to constrain all bond lengths and cut-off distances for the calculation of the coulombic and van der Waals interactions at 1.0 nm. The system was equilibrated by 100 ps of MD runs with position restraints on the protein to allow the relaxation of the solvent molecules at 300 K and normal pressure. The system was coupled to the external bath by the Berendsen thermostat [28] using a coupling time of 0.1 ps. The pressure was maintained by coupling to a reference pressure of 1 bar and V-rescaling [29] the temperature with a coupling constant of 0.1 ps. The final MD calculations were performed for 1 ns under the same conditions, except the position restraints were removed. The results were analyzed using the standard software provided by the GROMACS package.

In silico site-directed mutagenesis

In silico site-directed mutagenesis has been widely used to identify the critical residues responsible for binding to a ligand. The in silico mutations performed using the mutate_model command of MODELLER are listed in Table 1. Amino acid variations affecting the stability, structure and function of the protein were calculated by the SDM server (<http://mordred.bioc.cam.ac.uk/~sdm/sdm.php>). SDM uses a statistical potential energy function to predict the effect of amino acid polymorphisms on the stability of proteins. The percentage variations in solvent accessibility at positions of mutation along with the differential free energies of folding for mutants were readily generated.

Protein ligand docking studies

The chemical structures of monosaccharides were extracted from pubchem (<http://pubchem.ncbi.nlm.nih.gov>). The structures of all four carbohydrate ligands (glucose, galactose, *N*-acetyl neuraminic acid, *N*-acetyl galactosamine) included in binding to CTB were retrieved in the two-dimensional MDL/SDF format. Three-dimensional coordinates for these molecules were generated using the

CORINA [30] software. The molecules were then read into Discovery Studio 2.0 for further treatment of energy minimization for 100 steps with the CHARMM force field. Genetic Optimization for Ligand Docking (GOLD) version 4.1.1 (Cambridge Crystallographic Data Centre, Cambridge, UK) was used for docking 50 times in the standard default settings. The standard default settings (population size 100, selection pressure 1.1, niche size 2, migrate 10, crossover 95, number of operations 100,000, number of dockings 10) were adopted for GOLD docking. For ligand–protein binding, 10 docking conformations (poses) were tested and the best GOLD score selected for studies. GOLD uses a genetic algorithm to explore the full range of ligand conformational flexibility with partial flexibility of the protein [31]. The docking procedure consisted of three interrelated components: (a) identification of the binding site; (b) the use of a search algorithm to effectively sample the search space (the set of possible ligand positions and conformations on the protein surface); (c) the application of a scoring function. To estimate the protein–ligand complexes, the scoring function for the GOLD score was employed on the basis of four components: (a) protein–ligand hydrogen bond energy (external H-bond); (b) protein–ligand van der Waals energy (external vdw); (c) ligand internal van der Waals energy (internal vdw); (d) ligand intramolecular hydrogen bond energy (internal H-bond). The binding affinity between the protein and ligand was estimated using the consensus scoring function X-Score V2.1 [32]. The ligand that presented the greatest interaction with the protein was plotted using the program LIGPLOT [33]. Hydrogen-bonding interactions were double-checked with the software GETNEARES, which is available with the program DOCK [34].

Results and discussion

ctxB gene sequencing

The sequencing of representative *V. cholerae* O139 strains carrying either El Tor or Calcutta type CTX Φ prophages revealed variations in *ctxB* genes corresponding to amino acid positions 28, 34 and 68, and belonging to new genotypes (Table 1). Taking into account variations in *ctxB*, including the ones already reported, the translated amino acid sequences were used in protein structure generation and in silico mutational studies.

Molecular dynamics simulations

One of the 20 modeled structures of the CTB protein obtained by MODELLER9v6 that retained binding properties identical to the crystal structure along with the G factor

was selected, and the resultant trajectory of simulations was analyzed to study the physical properties of the protein. The time evolution of the root mean square deviation (RMSD) was computed for the simulated structure by taking the modeled protein as the initial structure (Fig. 1). Based on intrinsic dynamics, structural stability and the improved relaxation of the modeled protein, the energy of the structure and the radius of gyration were calculated (Fig. 2a, b). The energy and RMSD calculations for CTB demonstrated that the protein is not very flexible over the timescale of the MD simulations. The model obtained after rigorous refinement by means of MD and EM was subjected to mutational studies. We generated the variants of CTB corresponding to the reported classification of genotypes of *V. cholerae* [10] harboring the *ctxB* genes using the final model. The mutant structures were subjected to another round of EM and then used for docking.

Protein structure validation

To validate the homology-modeled CTB, a Ramachandran map was drawn and the structure was analyzed by PROCHECK, a well-known protein structure checking program. It was found that the phi/psi angles of 93.7% of the residues fell in the most favored regions, 6.3% of the residues fell in the additional allowed regions, 0% fell in the generously allowed regions, and none of the residues fell in the disallowed regions (Fig. 3). The overall PROCHECK G factor for the homology-modeled structure was -0.10 . A decrease in the overall G factor was observed after MD simulation. This indicates that an increase in the number of bad dihedral angles of the modeled structure had occurred, possibly due to the MD simulation causing an unfavorable dihedral angle, which allowed the protein to overcome high energy barriers. The final structure was

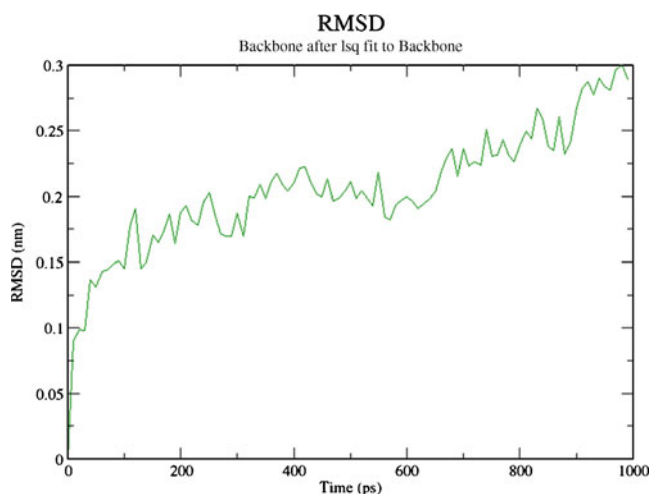


Fig. 1 Calculated root mean square deviation versus time graph for the modeled CTB protein

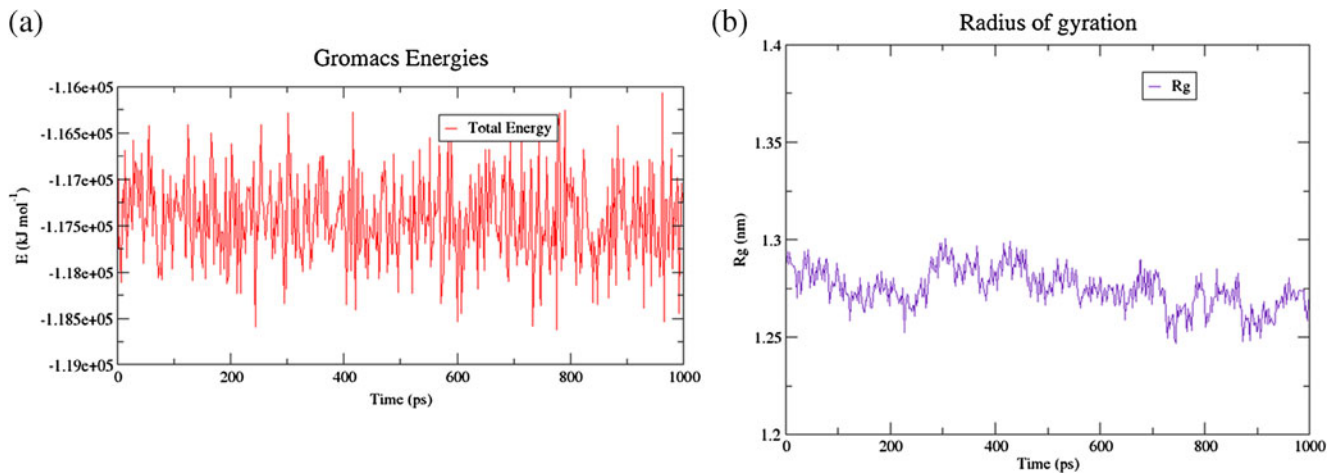


Fig. 2 **a** Calculated energy versus time plot for MD simulations of CTB using GROMACS software. **b** Radius of gyration over the timescale of MD simulations in CTB

validated by an ERRAT graph. The quality factor of 95.74 indicated good quality, as scores of >50 are acceptable for a reasonable model. These observations thus indicated that the obtained structure was of good quality.

Mutations and docking simulations

The translated protein sequences of the *ctxB* genotypes were utilized to perform protein model development, and the model used in silico mutagenesis. The structure of the CTB subunit predominantly consisted of β sheets sur-

rounded by α helices. The conserved Val 52–Ile 58 region encompassing a flexible loop plays an important role in the action of cholera toxin. We generated binding regions for the carbohydrates of GM1 pentasaccharide using all-atom superimposition of the template, which included His13, Asn14, Glu51, Gln56, His57, Gln61, Trp88, Asn90, and Lys91 [35]. The amino acids His13 and Asn14 fall in the loop region following a helix at the N-terminal region. Gln56 and His57 belong to the conserved flexible loop central to the action of the toxin, whereas Glu51 and Gln61 flank the loop region. Trp88, Asn90 and Lys91 were included in a β hairpin bend, and are known to be important in ligand binding. Though the mutations described here are not part of the flexible loop or the β hairpin bend (Fig. 4), we found differences in ligand binding among the variants of CTB, as presented below. Ligplots are provided in order to show variable hydrogen-bonding patterns among different CTB genotypes, which were considerably different from the template structure according to site-directed mutagenesis software. Protein–ligand interactions at energy minima representing the best docking poses were generated in accordance with the GOLD software. These complexes were used for ligplots that indicated probable hydrogen bonding patterns (indicated by dotted lines) in the respective variants (Fig. 5, 6, 7, 8). Mutations that had deleterious effects without being among the catalytic or substrate-binding residues were known to cause variations by triggering changes in structural features, such as alterations in surface charge allocation and the disruption of packing in the protein core region.

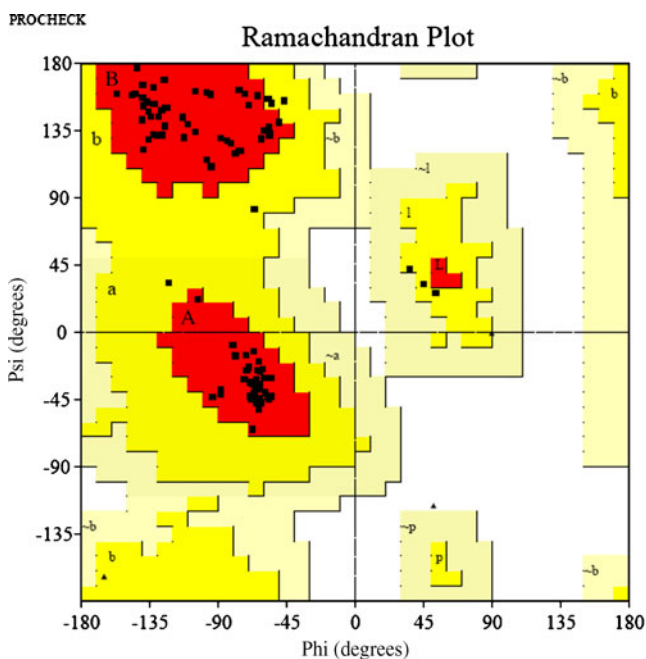


Fig. 3 Ramachandran map of the CTB protein. Calculations for the three-dimensional (3D) model of the CTB protein were done with the PROCHECK program

Genotype 1

ctxB genotype 1 has been reported in *V. cholerae* of the classical biotype carrying classical CTX Φ prophages. The

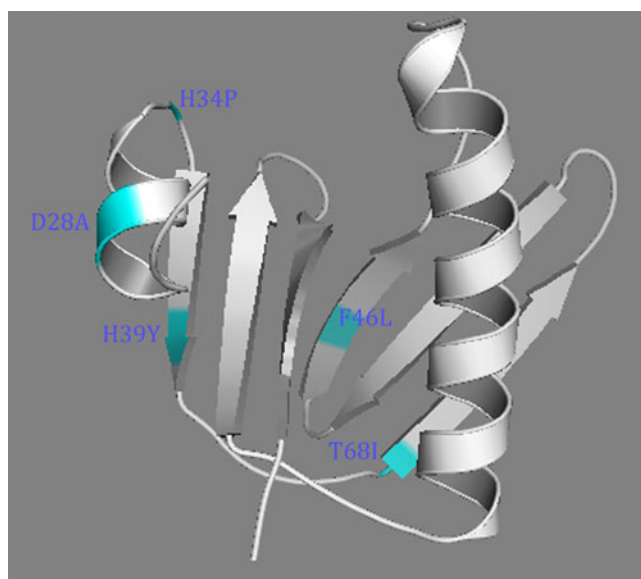


Fig. 4 CTB model labeled with mutations corresponding to different variants

amino acids at positions 39 (histidine), and 68 (threonine) are employed to classify the genotype of classical *ctxB*. Galactose, which is a major component of GM1 pentasaccharide, formed five hydrogen bonds in total, and five amino acids were involved in hydrophobic interactions within a radius of 10 Å from the binding pocket (Fig. 5a). Sialic acid and *N*-acetyl galactosamine each formed three hydrogen bonds with other amino acids coordinating hydrophobic interactions (Fig. 5b,c). Although glucose is a carbohydrate component of the GM1 receptor, no direct hydrogen bonding was seen among amino acids of the CTB. For all practical purposes, genotype1, corresponding to classical CT with the published crystal structure, will be

used as the standard against which the other types are compared.

Genotype 2

Genotype 2 showed the replacement of phenyl alanine with leucine at position 46, which occurs at the center of a β sheet. Docking simulations with galactose, sialic acid and *N*-acetyl galactosamine showed that in each case three hydrogen bonds formed between the ligand and the CTB protein, while 5–9 amino acids were involved in hydrophobic interactions. There were small increases in the percent solvent-accessible surface (+0.7) and the free energy of unfolding (+0.786 kcal mol⁻¹) compared to classical CTB for amino acid variation at this position. This variant does not show much of a difference in hydrogen bonding ability from classical CTB.

Genotype 3

Amino acid variations at positions 39 (histidine to tyrosine) and 68 (threonine to isoleucine) are the characteristic feature of genotype 3. Interestingly, docking simulations with galactose and *N*-acetyl galactosamine show only one hydrogen bond with each carbohydrate ligand (Fig. 6a,c), but four hydrogen bonds formed with sialic acid (Fig. 6b). The changes in solvent accessibility for amino acid substitutions at positions 39 and 68 for tyrosine and isoleucine were -7.9% and +28.9% compared to classical CTB. However, the free energy difference between the two proteins for unfolding was 3.040 kcal mol⁻¹. Therefore, this protein can be assumed to be physically more distinct than the other described genotypes.

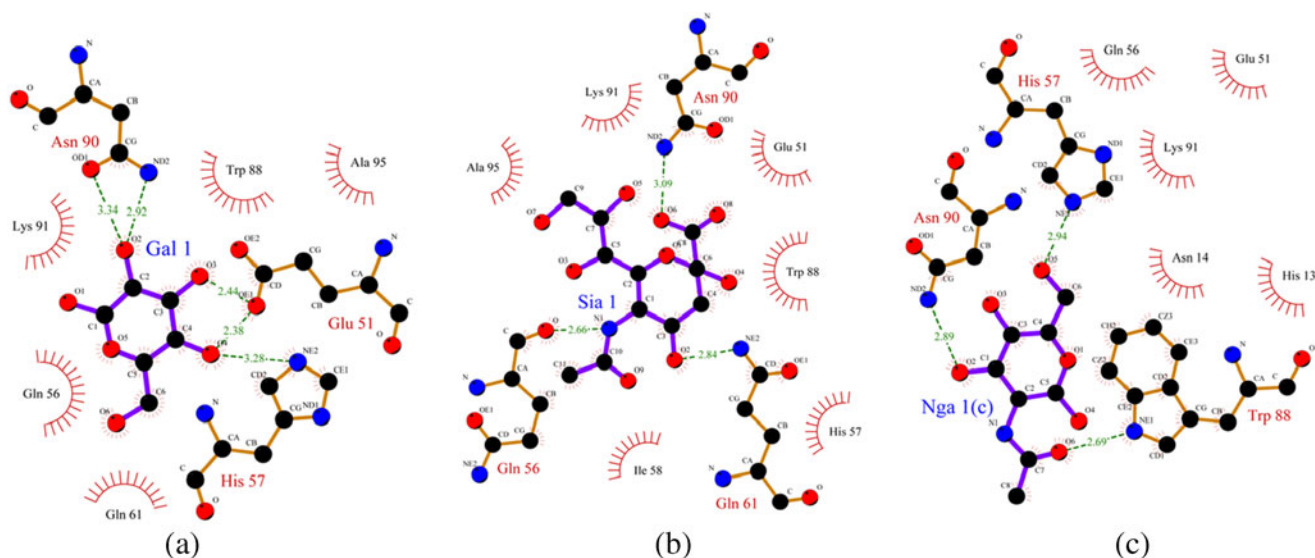


Fig. 5 Molecular interaction plots between carbohydrate ligands and genotype 1: **a** galactose, **b** sialic acid, **c** *N*-acetyl galactosamine

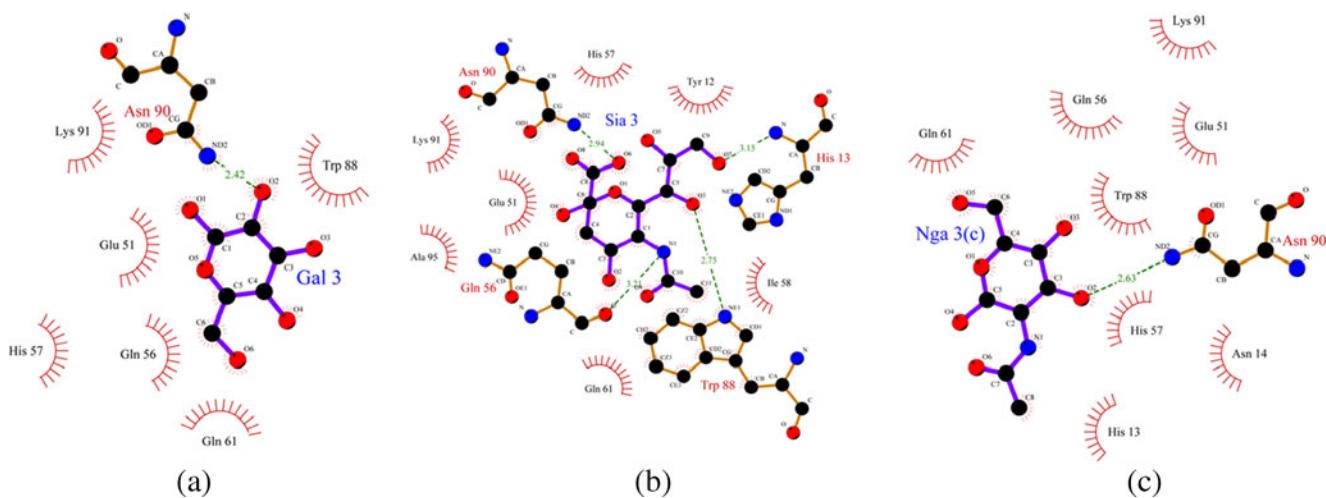


Fig. 6 Molecular interaction plots between carbohydrate ligands and genotype 3: **a** galactose, **b** sialic acid, **c** *N*-acetyl galactosamine

Genotype 4

Genotype 4 is characterized by an amino acid variation at position 39 from histidine to tyrosine. This variant formed three hydrogen bonds with each of the ligands employed, although the binding conformation was very different from that of genotype 1 at minimal energy. A decrease in the percent solvent accessible area (−4.7) along with a change in the differential free energy of unfolding of 1.82 kcal mol^{−1} was observed. However, this mutant was classified as neutral compared to classical CTB.

Genotype 5

This genotype was recently identified, and is associated with an amino acid change from aspartate to alanine at position 28. This CTB genotype also formed three

hydrogen bonds with all ligands except for glucose, which makes it similar to classical CTB in hydrogen-bonding ability (Fig. 7a–c). However, a decrease in the percent solvent accessible area (−9.1) was observed for this model compared to classical CTB. Also, the differential free energy of folding rose to 2.055 kcal mol^{−1}, thus distinguishing this variant of CTB.

Genotype 6

Genotype 6 was classified on the basis of variation at amino acid positions 34 (histidine to proline) and 39 (histidine to tyrosine) with respect to genotype 1. With galactose, this variant formed five hydrogen bonds, in a similar manner to classical CTB (Fig. 8a). The interactions between sialic acid and type 6 CTB were stronger with respect to hydrogen bonding (Fig. 8b). Aside from Asp90, Glu56 and Gln61 (as

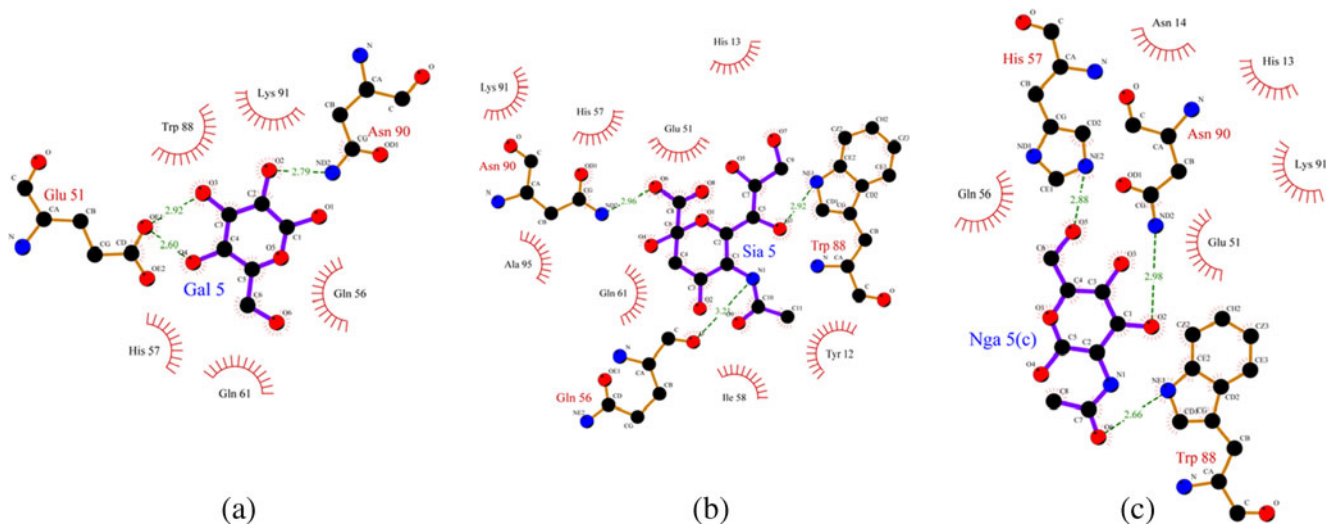


Fig. 7 Molecular interaction plots between carbohydrate ligands and genotype 5: **a** galactose, **b** sialic acid, **c** *N*-acetyl galactosamine

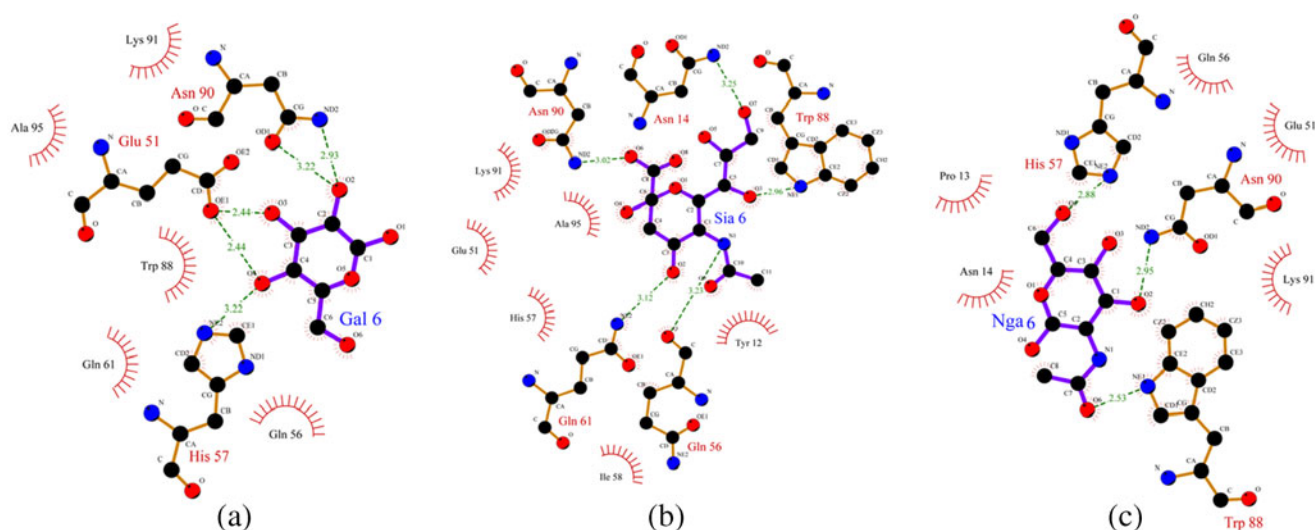


Fig. 8 Molecular interaction plots between carbohydrate ligands and genotype 6: **a** galactose, **b** sialic acid, **c** *N*-acetyl galactosamine

in genotype 1), Asn14 and Trp88 were also involved in direct hydrogen bonding between the protein and ligand. Although equivalent numbers of hydrogen bonds were observed for genotypes 1 and 6, their ligand binding characteristics with *N*-acetyl galactosamine differ, since about eight amino acids were involved in stabilizing the hydrophobic interactions, in contrast to classical CTB (Fig. 8c). Changes in the percent solvent accessibility at positions 34 and 39 were +21.8% and – 4.7% for this variant with respect to classical CTB. The free energy difference between the two proteins was 3.429 kcal mol⁻¹ for unfolding. These findings thus suggest that this CTB variant has a superior binding ability compared to other known CTB genotypes.

The classical type of CTB, designated genotype 1, was produced by the *V. cholerae* harboring the classical type of CTX Φ prophages. The El Tor type of CTX Φ prophages produces the El Tor type of cholera toxin, which includes two variants of CTB, designated genotypes 2 and 3. The O139 *V. cholerae* possesses both the El Tor and Calcutta types of CTX Φ prophages, and includes genotypic variants 4, 5 and 6. Sequencing data indicate that the majority of the *V. cholerae* O139 El Tor phages belong to genotype 4,

whereas the Calcutta-type phages are mostly of genotype 5. However, both variants of CTB produce the classical type of cholera toxin B (Fazil et al., unpublished data).

Docking simulations studies suggested that the ability of the type 1 CTB to interact with galactose and *N*-acetyl galactosamine is much greater than that of type 3 found in *V. cholerae* biotype El Tor (Table 2). Among the CTB variants, *V. cholerae* O139 genotype 6—the most recently discovered type—shows the most impressive bonding characteristics with galactose, sialic acid and *N*-acetyl galactosamine of all known CTB genotypes.

V. cholerae O1 biotype El Tor, the causative organism of the ongoing seventh pandemic, produces CT of genotype 3, while the classical biotype produces CT of genotype 1. The CT genotype of the El Tor strains currently associated with cholera in the Indian subcontinent has shifted from genotype 3 to genotype 1. Thus, the currently circulating El Tor strains that cause cholera have characteristics of the El Tor biotype but possess classical CTB [36, 37].

In terms of the chronology of evolution in *ctxB* genotypes, type 1 CTB was the first genotype of classical *V. cholerae* to be reported. The phage that encoded this CTB genotype was found to be integrated into host

Table 2 Variable docking scores of carbohydrate ligands with cholera toxin B subunit variants

Cholera toxin genotype	Docking score with respective carbohydrate ligand		
	Galactose	Sialic acid	Galactosamine
CT B1	32.51	37.56	28.52
CT B2	30.49	36.6	28.52
CT B3	24.53	37.15	25.32
CT B4	29.85	38.23	28.89
CT B5	30.35	38.15	28.83
CT B6	32.34	38.74	29.92

chromosomes and unable to replicate independent of the host, while the El Tor type phages replicated in the host to produce functional phages [38]. *V. cholerae* of the classical biotype are known to cause more severe cholera [39], while El Tor biotype strains survive better in aquatic environments [40]. The results of this study indicate decreased binding efficiency of El Tor CTB compared to the classical cholera toxin towards various carbohydrate ligands. The data generated suggest that new *ctxB* genotypes of *V. cholerae* O139, similar to classical CTB, possessed enhanced potential to bind carbohydrate ligands. Recent variations in CTB employed by the microbe indicate that *V. cholerae* produces a potent toxin that is more proficient at hydrogen bonding and thus forming stable complexes than the seventh pandemic clone of biotype El Tor. This could be a reason to worry, because current strains of *V. cholerae* O1 and O139 host to CTX prophages that have acquired the capacity for independent replication and infection, and could lead to the rapid spread of severe cholera.

Apart from the above mentioned ligands, we tested variant genotypes for their abilities to dock with glucose, which is in fact another carbohydrate component of monosialoganglioside. We found no interaction between the ligand and the proteins in any of the variant CTB proteins included in this study. In the case of CTB, it is understood that galactose, which is present at the GM1 terminus, has a potential role to play in interactions with CTB. Therefore, the elaboration of the binding characteristics of this specific carbohydrate would be primarily intended for virulence target initiatives or vaccine design [41]. Several studies have individually studied with galactose derivatives as an aid to the design of inhibitors against this toxin [42]. Hence, it can be concluded that potential derivatives of the carbohydrates used as CT inhibitors may employ the binding characteristics elucidated in this study.

Conclusions

The rate of change in the genetic profile of toxigenic *V. cholerae* has been a cause for concern due to their ability to cause epidemic and pandemic cholera. Among the various markers used to monitor genetic changes, variations in CTB have been studied recently. In this study, we explored the possibility of variable binding efficiencies of CTB variants to constituent carbohydrates of GM1 ganglioside, an important step in causing severe cholera. The results of this study indicate that there are subtle variations in the hydrogen bonding abilities of CTB variants toward the carbohydrate ligands that constitute the GM1 receptor. The data presented indicate the possibility of rapid genetic reassortments among O139 serogroup *V. cholerae*, thus

giving rise to multiple variants of CTB. Though there are no phenotypic differences among strains harboring CTB variants, the data provide clues about potential molecular discrepancies in the binding of CTB to its receptor. These differences can be employed in the design of paratopes that can determine the specificity of either a precise biotype or a standard anticholera toxin monoclonal antibody. As CTB is an attractive drug candidate for autoimmune diseases as well as differentiation therapy, the data presented here can be exploited in the design of drugs or vaccines.

Acknowledgments This work was supported in part by a grant from the Indian Council of Medical Research, New Delhi (Immuno. 18/11/17-ECD-I) to DVS and funds contributed by the Department of Biotechnology, New Delhi, to the Institute of Life Sciences. A senior Research Fellowship awarded by the Indian Council of Medical Research, New Delhi, India, to MHU Turabe Fazil is gratefully acknowledged. The authors thank the reviewers for their suggestions.

References

- Nusrin S, Khan GY, Bhuiyan NA, Ansaruzzaman M, Hossain MA, Safa A, Khan R, Faruque SM, Sack DA, Hamabata T, Takeda Y, Nair GB (2004) Diverse CTX phages among toxigenic *Vibrio cholerae* O1 and O139 strains isolated between 1994 and 2002 in an area where cholera is endemic in Bangladesh. *J Clin Microbiol* 42:5854–5856
- Waldor MK, Mekalanos JJ (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272:1910–1914
- Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406:477–483
- Davis BM, Moyer KE, Boyd EF, Waldor MK (2000) CTX prophages in classical biotype *Vibrio cholerae*: functional phage genes but dysfunctional phage genomes. *J Bacteriol* 182:6992–6998
- Freter R, O'Brien PC (1981) Role of chemotaxis in the association of motile bacteria with intestinal mucosa: fitness and virulence of nonchemotactic *Vibrio cholerae* mutants in infant mice. *Infect Immun* 34:222–233
- Mekalanos JJ (1985) Cholera toxin: genetic analysis, regulation, and role in pathogenesis. *Curr Top Microbiol Immunol* 118:97–118
- Holmgren J, Lonnroth I, Svennerholm L (1973) Tissue receptor for cholera exotoxin: postulated structure from studies with GM1 ganglioside and related glycolipids. *Infect Immun* 8:208–214
- Spangler BD (1992) Structure and function of cholera toxin and the related *Escherichia coli* heat-labile enterotoxin. *Microbiol Rev* 56:622–647
- Olsvik O, Wahlberg J, Petterson B, Uhlen M, Popovic T, Wachsmuth IK, Fields PI (1993) Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains. *J Clin Microbiol* 31:22–25

10. Bhuiyan NA, Nusrin S, Alam M, Morita M, Watanabe H, Ramamurthy T, Cravioto A, Nair GB (2009) Changing genotypes of cholera toxin (CT) of *Vibrio cholerae* O139 in Bangladesh and description of three new CT genotypes. *FEMS Immunol Med Microbiol* 57:136–141
11. Jertborn M, Nordstrom I, Kilander A, Czerkinsky C, Holmgren J (2001) Local and systemic immune responses to rectal administration of recombinant cholera toxin B subunit in humans. *Infect Immun* 69:4125–4128
12. Jertborn M, Svennerholm AM, Holmgren J (1994) Immunological memory after immunization with oral cholera B subunit-whole-cell vaccine in Swedish volunteers. *Vaccine* 12:1078–1082
13. Tamplin ML, Ahmed MK, Jalali R, Colwell RR (1989) Variation in epitopes of the B subunit of El Tor and classical biotype *Vibrio cholerae* O1 cholera toxin. *J Gen Microbiol* 135:1195–1200
14. Jobling MG, Holmes RK (2002) Mutational analysis of ganglioside GM (1)-binding ability, pentamer formation, and epitopes of cholera toxin B (CTB) subunits and CTB/heat-labile enterotoxin B subunit chimeras. *Infect Immun* 70:1260–1271
15. Aman AT, Fraser S, Merritt EA, Rodighiero C, Kenny M, Ahn M, Hol WG, Williams NA, Lencer WI, Hirst TR (2001) A mutant cholera toxin B subunit that binds GM1- ganglioside but lacks immunomodulatory or toxic activity. *Proc Natl Acad Sci USA* 98:8536–8541
16. Minke WE, Hong F, Verlinde CLMJ, Hol WGJ, Fan E (1999) Using a galactose library for exploration of a novel hydrophobic pocket in the receptor binding site of the *Escherichia coli* heat-labile enterotoxin. *J Biol Chem* 274:33469–33473
17. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1995) Short protocol in molecular biology, 3rd edn. Wiley, New York
18. Mantri CK, Mohapatra SS, Singh DV (2010) Effect of storage and sodium chloride on excision of CTX Φ or pre- CTX Φ and CTX Φ from *Vibrio cholerae* O139 strains. *Infect Genet Evol* 10:925–930. doi:10.1016/j.meegid.2010.05.015
19. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
20. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815
21. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291
22. Eisenberg D, Luthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol* 277:396–404
23. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J Comput Chem* 4:187–217
24. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 2:1511–1519
25. Hess B, Kutzner C, Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theor Comput* 4:435–447. doi:10.1021/ct700301q
26. Chris O, Alessandra V, Alan E, Wilfred FVG (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* 25:1656–1676
27. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472
28. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690
29. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126:014101
30. Tetko IV, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, Palyulin VA, Radchenko EV, Zefirov NS, Makarenko AS, Tanchuk VY, Prokopenko VV (2005) Virtual computational chemistry laboratory—design and description. *J Comput Aid Mol Des* 19:453–463
31. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 267:727–748
32. Wang R, Lai L, Wang S (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aid Mol Des* 16:11–26
33. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng* 8:127–134
34. Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule database. *J Comput Aid Mol Des* 15:411–428
35. Merritt EA, Sarfaty S, van den Akker F, L'Hoir C, Martial JA, Hol WG (1994) Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide. *Protein Sci* 3:166–175
36. Nair GB, Qadri F, Holmgren J, Svennerholm AM, Safa A, Bhuiyan NA, Ahmad QS, Faruque SM, Faruque AS, Takeda Y, Sack DA (2006) Cholera due to altered El Tor strains of *Vibrio cholerae* O1 in Bangladesh. *J Clin Microbiol* 44:4211–4213
37. Safa A, Sultana J, Dac Cam P, Mwansa JC, Kong RY (2008) *Vibrio cholerae* O1 hybrid El Tor strains, Asia and Africa. *Emerg Infect Dis* 14:987–988
38. Davis BM, Waldor MK (2003) Filamentous phages linked to virulence of *Vibrio cholerae*. *Curr Opin Microbiol* 6:35–42
39. Kaper JB, Morris JG Jr, Levine MM (1995) Cholera. *Clin Microbiol Rev* 8:48–86
40. Faruque SM, Albert MJ, Mekalanos JJ (1998) Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*. *Microbiol Mol Biol Rev* 62:1301–1314
41. Branderhorst HM, Liskamp RMJ, Visser GM, Pieters RJ (2007) Strong inhibition of cholera toxin binding by galactose dendrimers. *Chem Commun* 47:5043–5045. doi:10.1039/b711070g
42. Turnbull WB, Precious BL, Homans SW (2004) Dissecting the cholera toxin–ganglioside GM1 interaction by isothermal titration calorimetry. *J Am Chem Soc* 126:1047–1054

A comparative study of the hydrogen-bonding patterns and prototropism in solid 2-thiocytosine (potential antileukemic agent) and cytosine, as studied by ^1H - ^{14}N NQDR and QTAIM/ DFT

Jolanta N. Latosińska · Janez Seliger · Veselko Žagar · Dorota V. Burchardt

Received: 15 October 2010 / Accepted: 13 February 2011 / Published online: 29 March 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract A potential antileukemic and anticancer agent, 2-thiocytosine (2-TC), has been studied experimentally in the solid state by ^1H - ^{14}N NMR-NQR double resonance (NQDR) and theoretically by the quantum theory of atoms in molecules (QTAIM)/density functional theory (DFT). Eighteen resonance frequencies on ^{14}N were detected at 180 K and assigned to particular nitrogen sites ($-\text{NH}_2$, $-\text{N}=\text{}$, and $-\text{NH}-$) in 2-thiocytosine. Factors such as the nonequivalence of molecules (connected to the duplication of sites) and possible prototropic tautomerism (capable of modifying the type of site due to proton transfer) were taken into account during frequency assignment. The result of replacing oxygen with sulfur, which leads to changes in the intermolecular interaction pattern and molecular aggregation, is discussed. This study demonstrates the advan-

tages of combining NQDR and DFT to extract detailed information on the H-bonding properties of crystals with complex H-bonding networks. Solid-state properties were found to have a profound impact on the stabilities and reactivities of both compounds.

Keywords 2-Thiocytosine · Cytosine · ^{14}N -NQDR · DFT · QTAIM · Hydrogen bonding · Annular prototropism · Intermolecular interaction pattern · Molecular aggregations

Introduction

The compound called 2-thiocytosine (2-TC), an analog of the canonical natural nucleic acid base cytosine where the oxygen has been replaced with sulfur, is present in the tRNA of several organisms, including *Escherichia coli* [1]; however, it has not been detected in the DNA or RNA of natural mammals. The replacement of cytosine by 2-thiocytosine may cause significant changes in DNA structure due to the perturbation of the base-pairing process, or it can produce point mutations. Any alteration of or modification to the base-pairing scheme of DNA due to the existence of a different tautomeric form also may result in the perturbation of the replication process and spontaneous mutations (i.e., reduced stability of DNA). The presence of an abnormal tautomer of a thiosubstituted base can be even more deleterious to the stability of DNA. On the other hand, these unique abilities to modify DNA may be relevant in the design of DNA-binding drugs with high antitumor efficacies. Actually, 2-TC possesses important biological properties; it has shown to have a significant biological effect on the mitosis of human lymphocytes and

J. N. Latosińska (✉)
Faculty of Physics, Adam Mickiewicz University,
Umultowska 85,
61-614 Poznań, Poland
e-mail: Jolanta.Latosinska@amu.edu.pl

J. Seliger · V. Žagar
“Jozef Stefan” Institute,
Jamova 39,
1000 Ljubljana, Slovenia

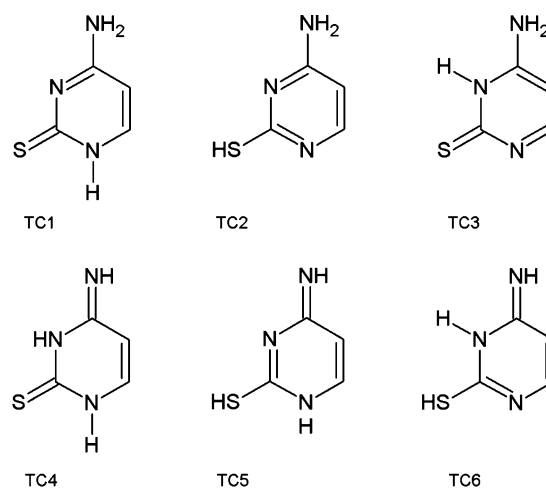
J. Seliger
Faculty of Mathematics and Physics, University of Ljubljana,
Jadranska 19,
1000 Ljubljana, Slovenia

D. V. Burchardt
Department of Paediatric Dentistry,
Karol Marcinkowski University of Medical Sciences,
Bukowska 70,
60-812 Poznań, Poland

is thus a potential antileukemic [2] and anticancer agent [3–9]. The antileukemic activity of fluorinated 2-TC has been proven only recently [10]. Numerous substituted 2-TC derivatives, including 1-(β -D-arabinofuranosyl)-2-thiocytosine and its analogs and complexes with trimethylplatinum have been synthesized, and their enzymatic reactivities and antitumor activities have been studied [11–13]. The cytotoxic activities of some of them were found to be even higher than that of cisplatin, and they were even active against cisplatin-resistant cell lines [13]. It was revealed that the nucleosides of 2-TC exhibit moderate inhibitory activity against Epstein–Barr viruses (EBV) in cell cultures [14, 15], and halogenated nucleosides of 2-TC exhibit considerable inhibitory activity against herpes simplex virus (HSV), and are potent inhibitors of varicella zoster virus (VZV) replication [16]. It was recently found that sulfur metabolic pathways are essential for survival and expression of virulence in many pathogenic bacteria, including *Mycobacterium tuberculosis* [17], so 2-TC (which contains sulfur) may represent a valuable lead for antibacterial and antituberculosis drug development.

Experimental [18–20] and theoretical [21–25] studies have shown that cytosine occurs in six tautomeric forms (the two most stable are the amino-oxo and amino-hydroxy forms), whose relative stabilities depend on the environment. It is well known that in the crystalline state, cytosine adopts its amino-oxo tautomeric form, with four planar molecules arranged in sheets exhibiting a network of H-bonds involving $-\text{NH}_2$, $-\text{NH}$, $-\text{N}=\text{C}$ and $\text{C}=\text{O}$ groups [24, 25]. 2-TC, just like cytosine, exists in six tautomeric forms (Fig. 1) as a result of thiol–thione and amino–imino equilibria. It was recently shown that its environment has an important impact on the tautomeric equilibrium of 2-TC (just as it does for cytosine); this equilibrium is significantly different in the gas phase (the predominant tautomer is the amino-thiol form), in solution (the predominant tautomer is the 1H-thione-amino form), and in nitrogen and argon matrices at low temperatures (exclusively amino-thiol) [26–29].

The preference for different tautomeric forms is a clear indication of the importance of intermolecular interactions, in particular H-bonding, in determining the structure of the condensed phase. The dominant tautomer in the crystalline state of 2-TC is—according to X-ray studies—the amino-thione form [30], but the channel required for proton migration (i.e., hydrogen bonds) is available in the crystalline structure. The arrangement of molecules in the crystals of 2-TC is essentially the same as that found for cytosine, but it is more complicated due to the existence of two nonequivalent molecules in the elementary cell. The presence of the “ordered” and “disordered” H-bonds in crystalline cytosine recently detected using the low-temperature FTIR [31] suggests that the possibility of



TC1>>TC3>>TC2>TC4>>TC6>TC5

TAUTOMER	ΔE [kJ/mol]	μ [D]
TC1	-	3.09
TC3	62.8	13.05
TC2	92.1	4.56
TC4	119.2	9.40
TC6	148.1	6.74
TC5	155.3	9.08

Fig. 1 Tautomeric structures of 2-thiocytosine

formal proton migrations in cytosine requires further investigation. In our previous papers, we showed that ^{14}N NQDR and QTAIM/DFT combined studies are extremely helpful for interpreting the complicated NQR spectra and for studies of intermolecular interactions in the crystalline state [25, 32–34].

^{14}N NDQR is the method of choice because ^{17}O NQR cannot be applied to 2-TC, although it is a great aid for identifying tautomers, as the ^{17}O parameters are very sensitive to oxygen hybridization and hydrogen bonds. It is worth noting that the electron density distribution at the $-\text{N}=\text{C}$, $-\text{NH}$ or $-\text{NH}_2$ sites is not only more accurately but it is also fully experimentally determined by NQDR. The same cannot be said of any structural method, such as standard X-ray diffraction (where spherically averaged pseudo-atom electronic distributions are used), because of the sensitivity of NQDR to short-range interactions, in contrast to the long-range periodic order seen by the X-ray-based techniques. Although the electron density distribution in the unit cell of a single crystal can be calculated from the structural factors determined in X-ray density quality studies, this electron density is loaded with error arising from experimental errors in structural factors and Fourier truncation errors (only a finite number of reflections can be collected), and phase information is lost during the measurement (the measured intensities are proportional to

the squared structural factors). Thus, density quality studies actually require further modeling of the electron density using independent atoms [35] or a multipole approach [36], and the result is not purely experimental.

Additionally, the study of weak interactions has recently evoked enormous interest, as they commonly occur in important biological systems. The fact that the less abundant tautomer is often the most reactive one (Hammond rule [37]) further reinforces the relevance of the use of combined ^{14}N NDQR and QTAIM/DFT methods to investigate intermolecular interactions in solid-state 2-TC, and to compare these interactions with those occurring in cytosine.

Experimental

A high-purity polycrystalline sample of 2-thiocytosine (97%) was purchased from Sigma–Aldrich and used without any additional purification.

^1H - ^{14}N -NQDR

NQR spectroscopy provides information on the quadrupolar interaction energy of the nuclear charge distribution with the electric field gradient (EFG) evaluated at the nuclear site.

When there is no magnetic field, a ^{14}N nucleus ($I=1$) has three generally nondegenerate nuclear quadrupole energy levels. The resonance (NQR) frequencies, usually termed $\nu_+ > \nu_- \geq \nu_0$, are expressed as [38]:

$$\begin{aligned} \nu_+ (^{14}\text{N}) &= \frac{e^2Qq}{4h} (3 + \eta) \\ \nu_- (^{14}\text{N}) &= \frac{e^2Qq}{4h} (3 - \eta) \\ \nu_0 (^{14}\text{N}) &= \nu_+ (^{14}\text{N}) - \nu_- (^{14}\text{N}) = \frac{e^2Qq}{2h} \eta \end{aligned} \quad (1)$$

where the $e^2Qq/h = e^2Qq_{zz}/h$ is the nuclear quadrupole coupling constant, $\eta = \left| \frac{q_{yy} - q_{xx}}{q_{zz}} \right|$ is the asymmetry parameter, $q_{ii} = \frac{\partial^2 V(r)}{\partial x_i^2}$ ($i=x, y$ and z) are the principal components of the second-rank symmetric EFG tensor, and $V(r)$ is the external electrostatic potential, which satisfies $|q_{xx}| \leq |q_{yy}| \leq |q_{zz}|$. The quadrupole coupling constant e^2Qq/h and the asymmetry parameter η for a ^{14}N nucleus can be derived from these frequencies using the formulae:

$$\begin{aligned} e^2Qq/h &= \frac{2}{3} [\nu_+ (^{14}\text{N}) + \nu_- (^{14}\text{N})] \\ \eta &= 2 \frac{[\nu_+ (^{14}\text{N}) - \nu_- (^{14}\text{N})]}{e^2Qq/h} \end{aligned} \quad (2)$$

Different double resonance techniques based on magnetic field cycling were used to detect the ^{14}N NQR frequencies.

The proton spin system was polarized in $B_0=0.75$ T for 30 s. Then the sample was pneumatically transferred within 0.1 s into another magnet, where it was left for 0.3 s. In this other magnet, the magnetic field can be varied continuously between zero and 0.1 T. After staying in this other magnet for <0.1 s, the sample was pneumatically transferred back into the first magnet, and the proton NMR signal was measured immediately after the sample had been stopped in the first magnet.

As the first method we used the ^1H - ^{14}N cross-relaxation spectroscopy [39, 40]. In this method, the sample is left to relax in a low magnetic field for 0.3 s, and the low magnetic field is changed between magnetic field cycles in steps of approximately 0.5 mT corresponding to the step in the proton Larmor frequency ν_L of 20 kHz. When the proton Larmor frequency ν_L matches the ^{14}N NQR frequency ν_Q , the proton spin-lattice relaxation time shortens, which results in a decrease in the proton NMR signal after the cycle. In some cases, especially at higher proton Larmor frequencies, a step of 40 kHz can be used. On the other hand, at around $\nu_L = \nu_Q$, the step is reduced to 10 kHz to improve the resolution.

In the second step, we used the solid-effect technique [41]. In this method, the low magnetic field is fixed and the sample is irradiated in the low magnetic field for 0.5 s with a strong rf magnetic field at variable frequencies. When the frequency ν of the rf magnetic field is equal to $\nu_Q \pm \nu_L$, simultaneous spin flips take place in both the ^1H and ^{14}N spin systems, so the proton magnetization drops to a lower value. This experiment is repeated at a few low magnetic field values in order to clarify the spectrum and get rid of the signal artefacts caused by direct proton absorption of the rf power at multiples of the proton Larmor frequency and the level crossing signals produced by the higher harmonics of the rf magnetic field.

As a final technique combining the three ^{14}N NQR frequencies assigned to a given nitrogen site, we used the two-frequency irradiation technique [42]. Here, the proton Larmor frequency ν_L is set in resonance with the lowest ^{14}N NQR frequency ν_0 and the sample is irradiated with two rf magnetic fields at the frequencies $\nu_1 = \nu$ and $\nu_2 = \nu + \nu_0$. When $\nu_1 = \nu_-$ and $\nu_2 = \nu_- + \nu_0 = \nu_+$, the proton relaxation rate in the low magnetic field increases, which causes the proton NMR signal at the end of the magnetic field cycle to drop to a low value.

The application of all three double-resonance techniques is necessary to unambiguously determine the triplets (ν_+ , ν_- and ν_0) of the ^{14}N NQR frequencies arising from various nonequivalent nitrogen positions in the crystal.

In the NQR spectrum of the compound studied, eighteen resonance lines were detected at 180 K, corresponding to six nonequivalent nitrogen positions in the crystal. The accuracy of NQR frequency determination was 10 kHz.

DFT/QTAIM calculations

Quantum chemical calculations were carried out within the GAUSSIAN03TM code [43] run on the CRAY supercomputer at the Poznań Supercomputer and Network Centre (PCSS). All calculations were performed within the density functional theory (DFT) with an exchange–correlation hybrid functional, B3LYP (three-parameter exchange functional of Becke B3 [44] combined with the Lee–Yang–Parr correlation functional LYP [45]), using an extended basis set with polarization and diffuse functions, 6-311++G**, which has been shown to produce highly accurate electron densities and electric field gradient (EFG) tensors. Satisfactory reproduction of the EFG tensors at ¹⁴N quadrupolar nuclei by DFT in the solid requires the use of the methodology first applied by us in 1997 [46], which was then improved and widely discussed in our previous papers, including our most recent [25, 32–34], but is only briefly summarized here. The calculations were carried out under the assumption of the presence of monomers and clusters. Formation of molecular clusters was divided into three steps. The first was the formation of six molecular clusters composed of ten molecules which assumed their X-ray atomic positions; the second was a partial optimization achieved using the Berny algorithm, during which only the positions of the hydrogen atoms were allowed to relax while those of all other atoms remained frozen. The third step involved the assumption that the proton positions for the central molecule were the most representative (i.e., they were insignificantly disturbed when PBCs: periodic boundary conditions—were ignored). The clusters constructed in this way—built from equivalent molecules of a certain tautomeric form—were used for further calculations of the NQR parameters and for topological analysis of the electron density within Bader’s quantum theory of atoms in molecules (QTAIM). The results of the EFG and QTAIM calculations are only discussed below for the central molecules in these clusters. Such a methodology ensures quite adequate results and is a very good compromise from the point of view of time/disk space requirements and the accuracy with which the electron density distribution can be determined. Moreover, according to the definition of the EFG tensor components, they decrease very quickly with increasing distance; e.g., towards the *zz* axis as x^2/r^5 . In view of the above, it can be assumed that the long-range effect on the reproduction of EFG tensor components is smaller than the effect that follows from taking into account the $\pi\cdots\pi$ stacking interactions (the latter does not exceed 5%). It should be noted that a valid reproduction of the EFG tensor and consequently the abovementioned NQR parameters ensures that a highly reliable electron density distribution is obtained. Thus, the NQR parameters—quadrupole coupling constants, asymmetry parameters,

and frequencies at all nitrogen atoms—were theoretically calculated at the B3LYP/6-311++G** level assuming that different molecular aggregations were formed by the intermolecular interactions depending on the tautomeric forms.

Further theoretical analysis of the intermolecular interactions was performed within QTAIM [47]. In this approach, the electron density $\rho(r)$ of a molecule treated as a scalar field can be examined by analyzing its gradient vector field. The values of $\rho(r)$ at the start and end points of a gradient path that follows the largest increase in $\rho(r)$ take the extreme points (maxima, saddle points, or minima in the electron density) and, depending on the nature of the extremum, they are termed nuclear attractor, bond, ring, and cage critical points, and denoted NACP, BCP, RCP and CCP, respectively. The kind of extremum can easily be determined with the help of a Hessian matrix consisting of nine second-order derivatives of $\rho(r)$. In its diagonalized form, the Hessian provides the three eigenvalues $\lambda_1, \lambda_2, \lambda_3$ (satisfying $\lambda_1 \leq \lambda_2 \leq \lambda_3$), which characterize the curvature of $\rho(r)$ along the three main curvature axes at the point r . Each critical point in $\rho(r)$ can be classified by a rank equal to the number of nonzero eigenvalues λ_i and a signature calculated as the algebraic sum of the signs of the eigenvalues λ_i of the Hessian. Maxima in the negative Laplacian of the electron density are indicative of local charge concentrations called valence shell charge concentrations (VSCC), and are classified as either bonded (located along bonds) or nonbonded (associated with lone pairs or unshared single electrons) maxima. In addition to the electron density and its Laplacian, the ellipticity of the bond, $\varepsilon = -\lambda_1/\lambda_2 - 1$, the total electron energy density at BCP (H_{BCP}) as well as its components—the local kinetic energy density (G_{BCP}) and the local potential energy density (V_{BCP}), as well as the hydrogen bonding energy according to Espinosa [48] were calculated.

Results and discussion

The ¹H-¹⁴N NQDR spectrum of 2-TC as obtained by the solid-effect technique at 180 K is presented in Fig. 2. The experimental NQR parameters are tabulated in Table 1.

The assignment of the frequencies to particular nitrogen sites is not trivial. Admittedly, all resonance signals were detected and assigned for cytosine [25, 49–51], but this task was made easier by having one molecule in the elementary cell. Although the presence of different kinds of ¹⁴N sites (–NH₂, –N=, and –NH–) in 2-TC facilitates this task, as the NQR parameters (the coupling constant and asymmetry parameter) are significantly different for such sites, the assignment of eighteen NQR frequencies to particular nitrogen sites in the 2-TC molecule is still difficult

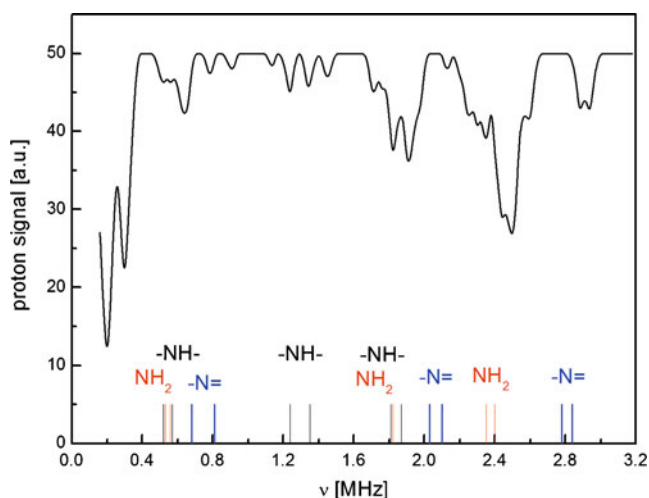


Fig. 2 The experimental ^1H - ^{14}N NQDR spectrum of 2-thiocytosine obtained at $T=180$ K by the solid effect technique at the proton Larmor frequency $\nu_L=100$ kHz. Using this technique, three dips are generally observed around the NQR frequency ν_Q at the frequencies ν_Q and $\nu_Q \pm \nu_L$. The dip at the frequency $\nu_Q + \nu_L$ is usually the most pronounced. The ^{14}N NQR frequencies, as confirmed by the two-frequency irradiation techniques and assigned to various nitrogen positions in the molecule on the basis of DFT calculations, are shown on the frequency scale

considering the nonequivalence of molecules (connected with the duplication of sites) and the possibility of prototropic tautomerism (capable of modifying the type of site due to proton transfer). The number of lines that can be grouped into six sets of three resonance lines (Fig. 2) with the help of the third equation of set (1) suggests that there are two tautomeric forms or two crystallographically nonequivalent molecules in the elementary cell. The latter seems to be consistent with the X-ray data [30], according to which 2-TC crystallizes in the monoclinic C2 space group with $a=19.5980(17)$, $b=4.0943(3)$, $c=13.5270(7)$ Å,

and $\beta = 97.54^\circ$. However, the presence of the “ordered” and “disordered” H-bonds in crystalline cytosine, which is structurally related, suggests that the possibility of formal proton migration in 2-TC should be considered.

The stability pattern of 2-TC tautomers in the gas phase ($\text{TC2} \gg \text{TC1} > \text{TC3} \gg \text{TC4} > \text{TC6} \gg \text{TC5}$), obtained at the B3LYP/6-311++G** level of theory, differs in the relative stability values but is generally in good agreement with the results obtained by Podolyan [9] at the MP2 and MP4 levels. To gain insight into the proton transfer processes in 2-TC in the solid, we modeled large clusters: decamers.

The stability pattern of 2-TC tautomers in the solid phase ($\text{TC1} > \text{TC3} \gg \text{TC2} > \text{TC4} \gg \text{TC6} > \text{TC5}$) differs from that in the gas phase, but only by the position of the tautomer (TC2) that is most stable in the gas phase. The DFT calculations predict that the TC2, TC4, TC5 and TC6 tautomers of 2-TC are much less stable than TC1 or TC3 in the solid state (see Fig. 1). The dipole moment of 2-thiocytosine (in TC1 form) is higher than that of the cluster (built from TC1 forms) (8.67 and 8.70 D versus 3.09 D).

Moreover, in the TC3 form, which is predicted to be energetically less stable than TC1, the monomer of 2-TC has much a higher dipole moment (10.2 D) and the cluster is also very polar (13.05 D). This suggests that in the solid (just as in polar media), intermolecular interactions stabilize the thione form. The dipole moment in general has a major impact on the packing properties in the solid state. Differences in the polarizations of adjacent molecules are closely connected with the specific arrangement. It can be expected that the polar nature of the molecules acts as a guiding template for a highly ordered solid-state structure. The predicted higher stability of TC1 than TC3 agrees well with the general observation that the requirement for a small number of transferred protons is an obstacle to prototropy

Table 1 The experimental NQR parameters for cytosine and 2-thiocytosine

Site	Cytosine					Reference	2-Thiocytosine					Molecule **
	e^2Qqh^{-1}	η	ν_+	ν_-	ν_0		e^2Qqh^{-1}	η	ν_+	ν_-	ν_0	
-NH ₂	2.916	0.394	2.49	1.941	0.548	[49] (RT)	2.780	0.381	2.35	1.82	0.53	A
	2.94	0.374	2.480	1.925	0.550	[50] (77 K)	2.813	0.398	2.40	1.82	0.56	B
	2.943	0.387	2.492	1.922	0.570	[51] (77 K)						
-NH-	2.180	0.699	2.016	1.247	0.769	[49] (RT)						
	2.17	0.757	2.035	1.215*	0.820	[50] (77 K)	2.147	0.484	1.87	1.35	0.52	A
	2.165	0.740	2.028	1.219	0.801	[51] (77 K)	2.033	0.561	1.81	1.24	0.57	B
-N=	2.865	0.786	2.712	1.586	1.126	[49] (RT)						
	***	–	–	–	–	[50] (77 K)	3.247	0.499	2.84	2.03	0.81	A
	2.862	0.806	2.723	1.570	1.153	[51] (77 K)	3.253	0.418	2.78	2.10	0.68	B
O	7.20	0.70	3.549*	2.002*	1.547*	[60]	–	–	–	–	–	–

* Not detected but calculated from the other data (frequencies or e^2Qq/h and η)

** Assignment given on the basis of ^{14}N NQDR and DFT results

*** Not detected

in infinitely long chains in solids [52]. Thus, the DFT method applied to solids produces reliable conclusions. To gain insight into the influence of proton transfer processes in the solid state on the NQR frequencies, e^2Qq/h and η at ^{14}N atoms in 2-TC, we performed the calculations under the assumption of the presence of six clusters built of ten molecules each at the B3LYP/6-311++G** level; see Table 2 and Fig. 3. A satisfactory reproduction of the e^2Qq/h , η and NQR frequencies at ^{14}N quadrupolar nuclei by DFT in solids requires the clustering methodology. The evident discrepancy between the experimental values and the results of DFT calculations performed using the monomer assumption (see Table 3 and Fig. 4) confirms the validity of our method. The quality of reproduction of the ^{14}N NQR parameters by DFT, assuming one of the most stable tautomers TC1 or TC3, is reasonable at first sight (the correlation coefficients are 0.992 and 0.977, and the standard deviations of the curve fit are 0.102 MHz and 0.339 MHz); see Fig. 4. However, the standard error of the curve fit for TC3 is three times as large as that for TC1. While the experimental NQR frequencies at the same type sites in both nonequivalent molecules differ by at most 0.11 MHz, the NQR frequencies calculated by DFT assuming TC1 differ by 0–0.08 MHz, while those calculated assuming TC3 differ by 0.03–0.55 MHz. The assumption of one of the four remaining tautomers TC2, TC4, TC5

and TC6 leads to unreliable values of e^2Qq/h (overestimated for TC2 and TC4, but underestimated for TC5 and TC6) and η (underestimated for TC2 and TC4 and overestimated for TC5 and TC6). This clearly shows that eighteen resonance lines detected by ^{14}N NQDR are the result of nonequivalence of molecules in the elementary cell, and, similarly to cytosine [25], only one tautomeric form predominates in the NQR spectra of 2-TC. It is worth noting that the stability pattern of 2-TC tautomers differs from that of cytosine [25], but in both only one tautomer predominates in the solid state. Thus, the presence of the “disordered” H-bonds in crystalline cytosine, as described by Rozenberg [31], results from the scattering of the proton position, but not from formal proton migrations. This conclusion is supported by the scatter in the experimental ^{14}N NQR data for cytosine (even at 77 K, which excludes the temperature factor) reported in the literature [49–51].

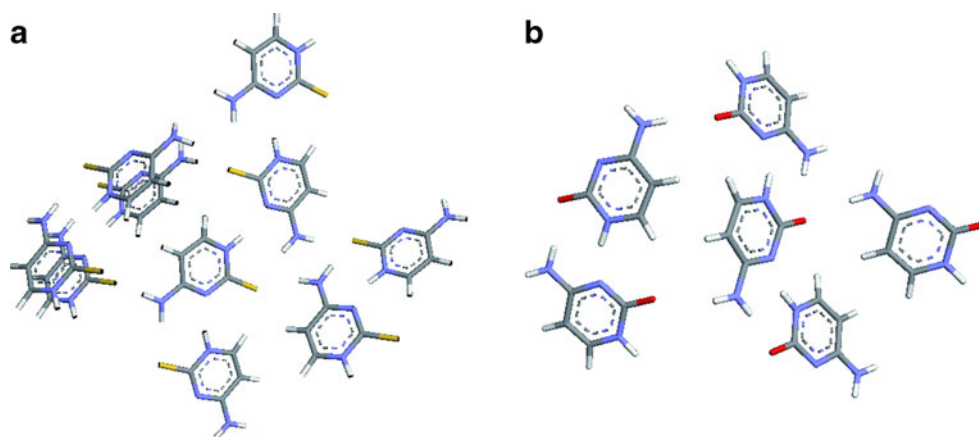
Pattern of intermolecular bonds

In the crystalline structure of 2-TC, the basic unit is a near-planar dimer made of two crystallographically nonequivalent molecules of the amino-thione tautomeric form, connected by two hydrogen bonds $\text{N}(1')\text{--H}(1')\dots\text{N}(3)$ and $\text{N}(4)\text{--H}(4)\dots\text{S}(2')$ with lengths of 3.022 and of 3.345 Å, respectively [30]; see Figs. 3a, 5. The structural unit formed

Table 2 NQR parameters calculated at the B3LYP/6-311++G(d,p) level of theory for 2-thiocytosine (clusters of ten molecules)

Tautomer	Formal site	Molecule A					Molecule B				
		e^2Qq/h	η	ν_+	ν_-	ν_0	e^2Qq/h	η	ν_+	ν_-	ν_0
TC1	–NH ₂	3.215	0.385	2.721	2.102	0.619	3.244	0.382	2.743	2.123	0.620
	–NH–	2.463	0.530	2.174	1.521	0.653	2.471	0.461	2.138	1.568	0.570
	–N=	3.684	0.518	3.240	2.286	0.954	3.649	0.541	3.230	2.243	0.987
	S=	37.301	0.521		20.338		37.386	0.540		20.510	
TC3	–NH ₂	3.279	0.466	2.841	2.077	0.764	3.535	0.413	3.016	2.286	0.730
	–NH–	2.687	0.315	2.227	1.804	0.423	2.341	0.586	2.099	1.413	0.686
	–N=	3.533	0.589	3.170	2.130	1.040	4.045	0.354	3.392	2.676	0.716
	S=	37.598	0.532		20.573		37.553	0.572		20.824	
TC2	–NH ₂	4.276	0.139	3.356	3.058	0.297	4.065	0.098	3.148	2.949	0.199
	–NH–	5.342	0.122	4.169	3.844	0.326	3.525	0.168	2.792	2.496	0.296
	–N=	4.836	0.177	3.841	3.413	0.428	3.758	0.225	3.030	2.607	0.423
TC4	–NH ₂	4.019	0.101	3.116	2.913	0.203	4.407	0.111	3.428	3.183	0.245
	–NH–	3.709	0.242	3.006	2.557	0.449	4.929	0.186	3.926	3.468	0.458
	–N=	3.416	0.175	2.711	2.413	0.299	5.299	0.110	4.120	3.829	0.291
TC6	–NH ₂	3.129	0.745	2.930	1.764	1.166	2.624	0.719	2.440	1.496	0.943
	–NH–	3.673	0.745	3.439	2.071	1.368	2.695	0.415	2.301	1.742	0.559
	–N=	2.873	0.344	2.402	1.908	0.494	4.073	0.545	3.610	2.500	1.110
TC5	–NH ₂	2.517	0.974	2.501	1.275	1.226	3.455	0.477	3.003	2.179	0.824
	–NH–	2.775	0.583	2.486	1.677	0.809	3.733	0.164	2.953	2.647	0.306
	–N=	3.644	0.513	3.200	2.266	0.935	2.694	0.418	2.302	1.739	0.563

Fig. 3 Molecular aggregations formed by the intermolecular bonds (1D layer): **a** 2-thiocytosine (12 molecules) and **b** cytosine (six molecules)



by the H-bonds, as shown in Fig. 5, can be termed a “supramolecular synthon” [53]. These supramolecular synthons (dimers) are linked to the neighboring supramolecular synthons by intermolecular interactions of the same pattern; i.e., weaker H-bonds: N(1)–H(1)...N(3') and N(4')–H(4')...S(2) of 3.114 and 3.408 Å. Neighboring ribbons are linked together by much longer (and thus weaker) N–H...S bonds with lengths of 3.466 and 3.551 Å. It is known that the following geometric criteria often hinder H-bonding: (1) a distance between the proton and the acceptor that is smaller than the sum of the van der Waals radii of these atoms; (2) a donor–proton–acceptor angle that is $>90^\circ$; (3) elongation of the proton–donor bond. Therefore, the so-called topological criteria formulated by Koch and Popelier [54] on the basis of Bader’s QTAIM theory are widely applied: (1) the presence of a BCP for the proton–acceptor contact as confirmation of the existence of a hydrogen-bonding interaction; (2) electron density at the proton–acceptor BCP that is within the range 0.002–0.040 au; (3) a corresponding Laplacian that is within the range 0.024–0.139 au; (4) “mutual penetration” of the hydrogen and acceptor atoms; (5) a loss of charge on the hydrogen atom; (6) energetic destabilization of the hydrogen atom; (7)

decreases in dipolar polarization and the volume of the hydrogen atom. The application of QTAIM not only allows us to verify the existence of these H-bond interactions, identify their nature and characterize their strengths, but it also permits the detection and distinction of many weaker interactions in the crystalline structure of 2-TC, which is not possible in standard X-ray studies [30]. As both N and S show relatively high electronegativities and have free electron pairs, under the right circumstances they can form multicenter bonds where one atom is simultaneously bonded to a few other atoms. Indeed, the crystalline structure of 2-TC contains four varieties of intermolecular interactions, such as the H-bonds N–H...N, N–H...S, C–H...S and C–H...C and three N...N, N...C and N...S interactions involving non-H atoms (see Fig. 5). Their molecular topology was described in terms of BCPs and RCPs, as shown in Table 4. Very small values of $\rho(r)$, small and positive values of the Laplacian, relatively high values of ϵ , near-zero values of H_{BCP} , and values of $|V_{BCP}|/G_{BCP} \leq 1$ indicate that these interactions are purely closed shell in the dichotomous classification based on the sign of the Laplacian [55, 56]; closed shell—excluding N(4)–H(4)...S(2'), which is of the transit type (it has partial covalent and

Table 3 NQR parameters calculated at the B3LYP/6-311++G(d,p) level of theory for single molecules of 2-thiocytosine (most stable tautomers)

Compound	Tautomer	Formal site	Molecule A					Molecule B				
			e^2Qq/h	η	ν_+	ν_-	ν_0	e^2Qq/h	η	ν_+	ν_-	ν_0
2-Thiocytosine	TC1	–NH ₂	4.024	0.293	3.31	2.72	0.59	4.060	0.247	3.300	2.790	0.510
		–NH–	2.492	0.804	2.37	1.37	1.00	2.643	0.494	2.310	1.660	0.650
		–N=	4.219	0.343	3.53	2.80	0.72	4.118	0.355	3.450	2.720	0.730
		S=	36.993	0.982	24.442	37.047	0.979	24.441				
	TC3	–NH ₂	5.008	0.070	3.844	3.668	0.175	5.045	0.065	3.866	3.702	0.164
		–NH–	3.479	0.300	2.870	2.348	0.522	3.491	0.297	2.877	2.359	0.518
		–N=	4.674	0.194	3.732	3.279	0.453	4.713	0.183	3.750	3.319	0.431
		S=	36.729	0.983		24.280		36.890	0.982		24.374	

The nonequivalent molecules A and B are depicted in Fig. 4

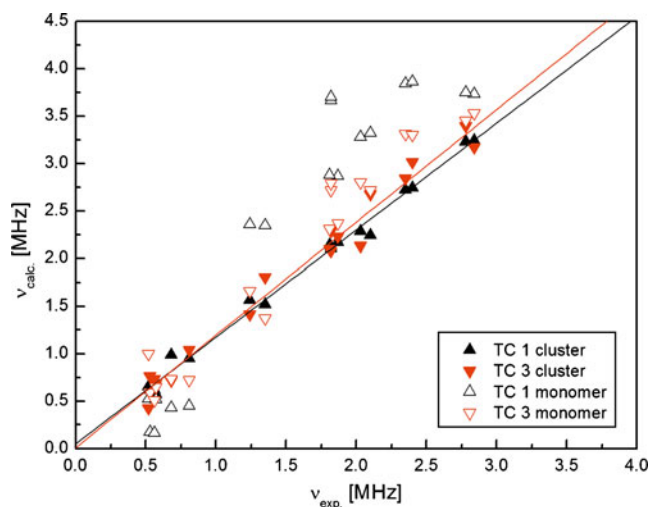


Fig. 4 The correlation between the experimental and calculated NQR frequencies (monomers and clusters of TC1 and TC3; *solid line* linear fit for cluster)

partial ionic character)—in terms of the classification based on the nondimensional $|V_{\text{BCP}}/G_{\text{BCP}}| < 1$ ratio [55, 57]; or closed shell—excluding N(4)–H(4)...S(2'), which is open shell—in terms of the classification based on the atomic valence shell and both the local (BCP) and integral properties [55, 58]. It is worth noting that for N(4)–H(4)...S(2') bonds, the H...S distance exceeds 2.2 Å, so their classification as weak H-bonds coincides with the classification proposed by Jeffrey [59], who considered that hydrogen bonds with $d(\text{H...Y}) > 2.2$ Å are weak and mainly electrostatic, and hydrogen bonds with $d(\text{H...Y}) < 1.5$ Å are strong and covalent.

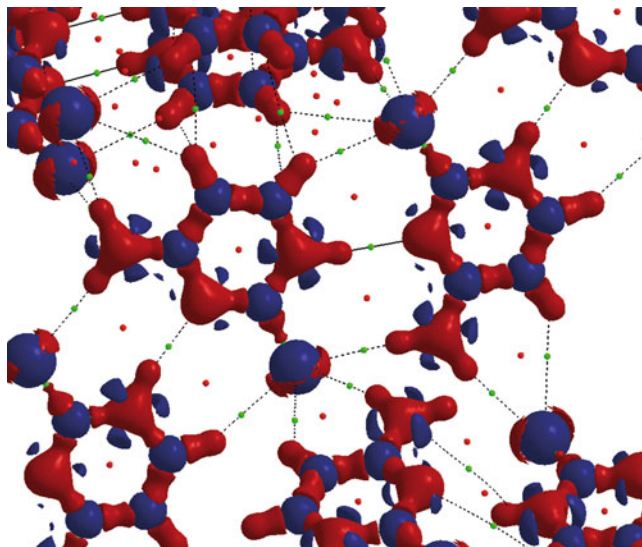


Fig. 5 The 3d distribution of the electron density Laplacian calculated by DFT for 2-thiocytosine (isocontour ± 0.35 a.u.); the regions of negative Laplacian are shown in red, the regions of positive Laplacian in blue, *small circles* correspond to critical points (*red* RCP, *green* BCP)

In terms of the total energy density, the N(1')–H(1')...N(3) interaction is found to be the strongest (-23.18 kJ mol $^{-1}$), followed by N(4)–H(4)...S(2') (-17.13 kJ mol $^{-1}$); see Table 4. The third H-bond linking adjacent molecules into a dimer, C(6)–H(6)...S(2'') (length: 3.532 Å, strength: -9.43 kJ mol $^{-1}$), is much weaker and is therefore not revealed by X-ray studies [30]. The hydrogen bonds linking adjacent supramolecular synthons are N(1)–H(1)...N(3') (3.114 Å, -16.10 kJ mol $^{-1}$) and N(4')–H(4')...S(2) (3.408 Å, -13.95 kJ mol $^{-1}$). The H-bonds linking neighboring ribbons are two N(4)–H(4)...S(2') (3.466 Å, -3.17 kJ mol $^{-1}$ and 3.551 Å, -8.01 kJ mol $^{-1}$), C(5)–H(5)...S(2') (4.037 Å, -2.85 kJ mol $^{-1}$), and two C(5)–H(5)...C(6') and C(5')–H(5')...C(6) (3.688, -3.99 kJ mol $^{-1}$ and 3.943 Å, -1.91 kJ mol $^{-1}$), respectively. The reasons for the differences in the strengths of the N–H...S bonds (each linking $-\text{NH}_2$ and S) are their different lengths and the lattice vibrations comprising the motions of the NH_2 groups, which produce averaging of the electron density and electric field gradient, which was not considered.

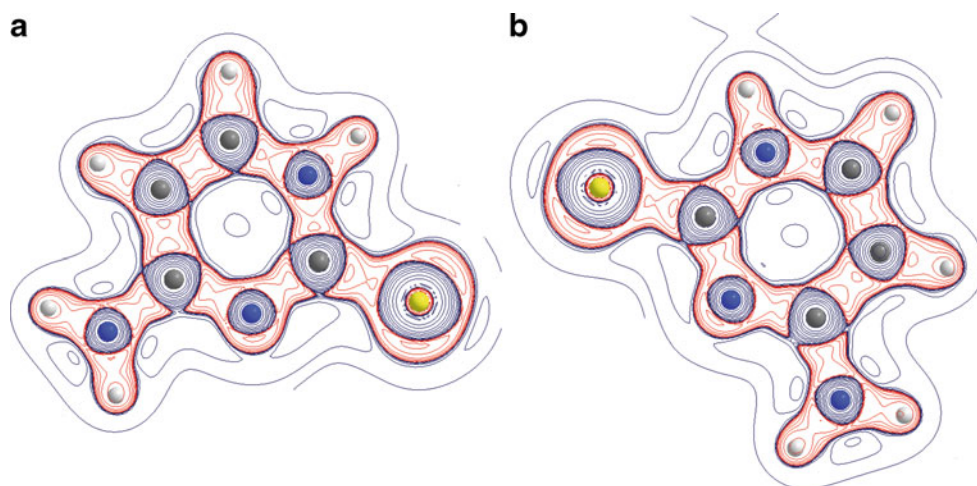
Additionally, a few π – π interactions (stacking) involving non-H atoms, purely van der Waals in nature, were detected between layered dimers: N(1)...N(3''), N(1)...C(5'') and N(3)...S(2''), with lengths of 4.094, 3.411 and 3.516 Å and approximate estimated energies of -1.14 , -4.06 and -4.01 kJ mol $^{-1}$. It is worth noting that N(4) from the NH_2 group opposite to N(1) and N(3) does not participate in these stacking π – π interactions. The hierarchy of structures in crystals of 2-TC (molecules, supramolecular synthons (dimers), ribbons, stacks) is reflected in progressively weaker bonds.

A more accurate analysis of the Laplacian contour maps and isosurfaces for 2-TC reveals additional details (see Figs. 5 and 6), which are important from the point of view of further comparison with cytosine. It is worth noting that three VSCCs that are oriented towards their bonding partners and consistent with sp^2 hybridization are found at the sulfur atom. Nonetheless, the presence of bond paths that pass through regions of low charge density and BCP clearly indicates the existence of five intermolecular bonds in addition to the C=S bond; see Figs. 5 and 6. It should be noted that three of these intermolecular bonds are in fact very weak, which may explain why they were difficult to discern. One of the VSCCs at the sulfur atom points towards the carbon atom, and one VSCC at the carbon points towards the sulfur atom. This suggests that the C=S bond is not strongly polarized. The other two VSCCs point towards the proton atoms from NH_2 groups bonded to sulfur via the N–H...S hydrogen bonds. In the isosurface representation of the Laplacian (see Fig. 5), the maxima corresponding to these VSCCs are pronounced and quite symmetric, irrespective of whether molecule A or B is analyzed. The search for other minima in the Laplacian around the $-\text{NH}_2$ identifies an additional VSCC, oriented

Table 4 Topological parameters of ρ for 2-thiothytosine and cytosine: the electron density at BCP and RCP $\rho_{\text{BCP}}(r)$, its Laplacian $\Delta_{\text{BCP}}(\rho)$, the potential electron energy density V_{BCP} , the kinetic electron energy density G_{BCP} , the total electron energy density H_{BCP} and the estimated hydrogen-bonding energy according to Espinosa (E_E) calculated at the B3LYP/6-311++G(d,p) level of theory

Compound	Critical point type	$R(X-H\dots Y)$ in Å	$R(Y\dots H)$ in Å	$\rho_{\text{BCP}}(r)$ in a.u.	$\Delta_{\text{BCP}}(\rho)$ in a.u.	ϵ	G_{BCP} in a.u.	V_{BCP} in a.u.	H_{BCP} in a.u.	$ V_{\text{BCP}} /G_{\text{BCP}}$	$G_{\text{BCP}}/\rho(r)$ in a.u.	$H_{\text{BCP}}/\rho(r)$ in a.u.	E_E in kJ/mol
2-Thiothytosine	N(1')-H(1')...N(3)	3.022	2.004	0.0272	0.0761	0.0622	0.0183	-0.0177	0.0006	0.97	0.67	0.02	-23.18
	N(4)-H(4)...S(2')	3.345	2.330	0.0235	0.0495	0.0890	0.0127	-0.0130	-0.0003	1.02	0.54	-0.01	-17.13
	C(6)-H(6)...S(2'')	3.532	2.581	0.0148	0.0384	0.0767	0.0084	-0.0072	0.0012	0.86	0.57	0.08	-9.43
	N(1)-H(1)...N(3')	3.114	2.108	0.0213	0.0629	0.0640	0.0140	-0.0123	0.0017	0.88	0.66	0.08	-16.1
	N(4')-H(4')...S(2)	3.408	2.382	0.0209	0.0444	0.0622	0.0109	-0.0106	0.0003	0.97	0.52	0.01	-13.95
	C(5)-H(5)...S(2')	4.037	3.165	0.0050	0.0137	0.3965	0.0028	-0.0022	0.0006	0.79	0.56	0.12	-2.85
	C(5)-H(5)...C(6)	3.688	2.390	0.0063	0.0204	0.8972	0.0040	-0.0030	0.0010	0.75	0.63	0.16	-3.99
	C(5')-H(5')...C(6)	3.943	3.102	0.0065	0.0082	0.8613	0.0016	-0.0012	0.0004	0.75	0.25	0.06	-1.91
	N(4)-H(4)...S(2')	3.466	2.458	0.0166	0.0400	0.0482	0.0090	-0.0081	0.0009	0.90	0.54	0.05	-3.17
	N(4')-H(4')...S(2'')	3.551	2.564	0.0128	0.0340	0.5120	0.0072	-0.0061	0.0011	0.85	0.56	0.09	-8.01
	N...S	3.517	-	0.0063	0.0192	4.5466	0.0039	-0.0031	0.0008	0.79	0.62	0.13	-4.06
	N...N	4.094	-	0.0018	0.0062	0.2816	0.0012	-0.0009	0.0003	0.75	0.67	0.17	-1.14
	N...C	3.411	-	0.0059	0.0161	0.6576	0.0035	-0.0031	0.0004	0.89	0.59	0.07	-4.01
	Cytosine	N(1')-H(1')...N(3)	2.84	1.89	0.034	0.102	0.084	0.026	-0.027	-0.0010	1.04	0.76	-0.03
N(4)-H(4)...O(2')		2.98	2.13	0.016	0.061	0.023	0.013	-0.011	0.0020	0.85	0.81	0.13	-14.08
C(5)-H(5)...O(2')		3.58	2.88	0.005	0.015	0.079	0.003	-0.003	0.0000	1.00	0.60	0.00	-3.30
N(4)-H(4)...O(2'')		3.03	2.15	0.016	0.067	0.086	0.014	-0.011	0.0030	0.79	0.88	0.19	-14.13
N...O		3.668	-	0.004	0.014	0.580	0.006	-0.003	0.0030	0.01	1.50	0.75	-3.35
N...N		3.814	-	0.003	0.008	0.621	0.002	-0.001	0.0010	0.50	0.67	0.33	-1.85
N...C		3.392	-	0.005	0.015	0.600	0.003	-0.003	0.0000	1.00	0.60	0.00	-3.50

Fig. 6 Laplacian contours in 2-thiocytosine. *Left:* molecule A; *right:* molecule B (negative regions in red, positive regions in blue)



towards the carbon atom. The search for minima in the Laplacian around each of the nitrogen atoms ($-\text{N}=\text{}$ and $-\text{NH}-$) of the ring gives three VSCCs that are oriented towards their bonding partners and consistent with sp^2 hybridization. The isosurface representation of the Laplacian around $-\text{N}=\text{}$ reveals the expected VSCC of the H-bond, which is oriented in the plane of the lone pair with a pronounced maximum; by contrast, the donor nitrogen atom $-\text{NH}-$ exhibits a more symmetrical distribution of the Laplacian and a VSCC that is directed towards H with an even more pronounced (higher) maximum. The VSCCs of carbon C(2) that are directed towards these nitrogen atoms are shifted to the nitrogen basins in both molecules. Nevertheless, the N–C and N–H bonds show features that are typical of covalent interactions. The differences between the isosurface representations of the Laplacian at the nitrogen sites in both nonequivalent molecules are small, but they suggest that the $-\text{NH}-$ site has a more symmetrical electron density distribution in molecule A than in B, while the $-\text{N}=\text{}$ and $-\text{NH}_2$ sites have more symmetrical electron density distributions in molecule B than in A, which is in a good agreement with experimentally obtained values.

2-Thiocytosine versus cytosine

Although the crystal structures of cytosine and 2-TC are generally similar, cytosine crystallizes with only one molecule in the asymmetric unit [24], which gives an extra advantage when studying its intermolecular interactions. A comparison of the geometry of single molecules of 2-TC and cytosine in the crystalline state reveals significant differences in the neighborhood of the C(2) atom, where oxygen is replaced with sulfur. The C(2)–S(2) bond in 2-TC is much longer than the C(2)–O(2) bond in cytosine (1.702 in molecule A and 1.701 in molecule B versus 1.24 Å). The absolute values of the potential electron energy density and the kinetic electron energy density are much smaller in 2-TC than in cytosine (–0.3983 for molecule A

and –0.4012 for molecule B versus –1.1657 a.u. and 0.1556 for molecule A and 0.1557 for molecule B versus 0.5237 a.u., respectively). The number of electrons shared by adjacent atoms, as measured by the delocalization index, at the BCP in 2-TC is higher than that in cytosine (1.372 for molecule A and 1.360 for molecule B versus 1.132), and the asymmetry ε describing the curvature of $\rho(r)$ and thus indicating multiple bonding or electronic depletion at the BCP is much lower for C=O than for C=S (0.0891 versus 0.1244 for molecule A and 0.1109 for molecule B). Therefore, we can conclude that C=S has more double-bond character than C=O.

However, the NBO analysis clearly shows that the Wiberg and the atom–atom overlap-weighted NAO bond orders for C=O are noticeably higher than for C=S (1.6073 versus 1.4941 for molecule A and 1.4975 for molecule B, and 1.226 versus 1.1420 for molecule A and 1.1451 for molecule B, respectively, for the Wiberg and NAO bond orders).

Detailed analysis of NBO orbitals suggests that C=O has more π and σ character than C=S, as revealed by the lower occupancies of σ and π for C=S than for C=O (σ : 1.9753 for molecule A and 1.9754 for molecule B versus 1.9926; π : 1.9685 for molecule A and 1.9692 for molecule B versus 1.9861). These NBO results are in good agreement with the sum of the Hessian eigenvalues $\lambda_1 + \lambda_2$ [–1.9143 for C=O and –0.4467 (for molecule A) and –0.4460 (for molecule B) for C=S] and with λ_3 (1.4404 versus 0.0981 for molecule A and 0.1027 for molecule B).

It is worth noting that $|\lambda_1/\lambda_3|$ is much lower for C=O than C=S (0.6928 versus 2.411 for molecule A and 2.287 for molecule B), which suggests that the C=O bond has more ionic character but the C=S bond is more covalent. The larger integrated net atomic charges in C=O than in C=S (–1.2221 at carbon and 1.7231 at oxygen in cytosine versus –0.1944 and 0.6875 for molecule A and –0.2153 and 0.6864 for molecule B at the carbon and sulfur, respectively, in 2-TC) agree with stronger polarization of

the C=O bond. The dipole moment of the bond, an indicator of polarizability, is much smaller for the C=S bond than the C=O bond (0.118 versus 1.128 D). The position of the BCP, which is located near the middle of the bond for C=O but is shifted to the sulfur atom for C=S, also complies with the higher polarizability of C=O.

Thus, the greater ellipticity of C=S than C=O (confirmed by the delocalization index value), which in fact varies along the bond path, does not seem to be the result of the higher bond order but rather the more symmetric BCP location, in view of the much smaller charge transfer in C=S than in C=O.

The adjacent bonds C(2)–N(3) and N(1)–C(2) in 2-TC are shorter than in cytosine (1.368 and 1.342 Å, versus 1.374 and 1.364 Å). The electron densities at the BCP and RCP in cytosine are lower than in 2-TC (BCP: 0.3297 e for C(2)–N(3) and 0.3095 e for N(1)–C(2) versus 0.3411 e for C(2)–N(3) and 0.3149 e for N(1)–C(2); RCP: 0.02392 versus 0.02413 e). λ_3 for C(2)–N(3) is higher than that for the N(1)–C(2) bond in cytosine (0.3757 versus 0.3677), in contrast to the situation in 2-TC, in which λ_3 is higher for N(1)–C(2) than for the C(2)–N(3) bond (0.3937 versus 0.3619). At the same time, the hybridization of the C(2)–N(3) and N(1)–C(2) bonds in cytosine and 2-TC differs insignificantly (by less than 5%). Moreover, the integrated net atomic charges at N(1), N(3), C(4), N(4) and H(5) increase, while those at C(2), C(5), C(6), H(1), H(4) and H(6) decrease upon substituting the sulfur atom with oxygen. Thus, the change in the curvature along the C(2)–N(3) and N(1)–C(2) bond paths reflects the charge polarization of the C–N bond due to the substitution of sulfur for oxygen, rather than a change in hybridization. The replacement of oxygen with sulfur means that there is no strong inductive effect of oxygen. The resonance donation effect (lone pairs on the oxygen adjacent to the π system), which increases the electron density on the ring and subsequently at the N(1) and N(3) sites *ortho* and decreases this density at C(2) adjacent to oxygen, exceeds the inductive electron-withdrawing effect (through σ bonds) of oxygen. This is because sulfur is less efficient than oxygen at donating its nonbonded (lone) electron pairs in a resonance interaction with the heterocyclic ring. These changes can be interpreted in terms of the flow of bonding electrons to the neighboring C–N bonds (especially the N(3)–C(2) bond) from the external double bond upon substituting the oxygen atom for sulfur. Again, these results are in good agreement with the results of NBO analysis, as the occupancy of N(3)–C(2) exceeds that of C(2)–N(1) in cytosine (1.9951 versus 1.98246), while for 2-TC the situation is reversed, as the occupancy of C(2)–N(3) is lower than that of C(2)–N(1) (1.98149 versus 1.98833).

Topological analysis of the electron density distribution using its Laplacian indicates that three VSCCs are found at

the oxygen atom. One of the VSCCs at the oxygen atom points towards the carbon atom, but there is practically no VSCC at the carbon pointing towards the oxygen atom, in contrast to the situation for the C=S in 2-TC, where there is a pronounced VSCC at the carbon pointing towards the sulfur atom. This justifies the description of the C=O bond as very strongly polarized. The presence of bond paths that pierce through the regions of low charge density clearly indicates that there are four intermolecular bonds (two C–H...O and two N–H...O) in addition to the C=O bond.

It should be noted that two of these intermolecular bonds of C–H...O type are in fact very weak, and the bonds of both types occur in close proximity, which explains why it is difficult to find VSCCs, as they are not very pronounced in the isosurface representation of the Laplacian. It is worth noting that the bond paths correspond to preferable interactions. A search for other minima in the Laplacian around each of the nitrogen atoms ($-\text{N}=\text{}$, $-\text{NH}-$ and $-\text{NH}_2$) in cytosine reveals that (just as in 2-TC) they form VSCCs that are oriented towards their bonding partners. The electron densities at almost every BCP (excluding N(3)–C(4) and C(4)–NH₂) in cytosine are higher than those in 2-TC.

The charge concentrations around the nitrogen $-\text{NH}-$ in cytosine and in 2-TC are much greater in the direction of the hydrogen atom than in the direction of the aromatic ring.

The reverse situation is found for $-\text{N}=\text{}$ and $-\text{NH}_2$; the most distinct concentration is oriented towards the aromatic ring and the less pronounced VSCCs into the hydrogen bond.

This suggests that the π electrons flow from C=O to not only neighboring bonds but also the whole heterocyclic ring. The differences between the isosurface representations of the Laplacian in cytosine and 2-TC are well pronounced at oxygen and sulfur, but the nitrogen $-\text{NH}-$ exhibits a more symmetrical distribution in 2-TC than cytosine, as reflected by an increase in the NQR asymmetry parameter. Detailed analysis of the changes in ρ_{BCP} suggests that the most pronounced changes in electron density are those at the N–H bonds of $-\text{NH}_2$ and at H-bonds created using the lone pair of $-\text{N}=\text{}$. Nonetheless, the NQR parameters at $-\text{NH}_2$ differ only slightly, while the change caused by replacing oxygen with sulfur is evident at $-\text{N}=\text{}$. Taking into account the orientations of the components of the EFG tensor at each nitrogen site and symmetry considerations, we can conclude that the *zz* axis of the EFG tensor is parallel to the bond path with the BCP at which there is a large change in electron density only at $-\text{N}=\text{}$. This explains why the highest decrease in the quadrupole coupling constant upon replacing sulfur with oxygen is observed at the $-\text{N}=\text{}$ site.

The structural differences caused by exchanging sulfur for oxygen are not limited to single molecules; this substitution also strongly influences the intermolecular

interaction pattern. One of the critical differences between oxygen and sulfur is sulfur's lower electronegativity, which is no less important than its larger van der Waals radius (180 pm instead of 152 pm). Sulfur is able to make less polar and longer bonds with hydrogen than oxygen, so it is expected to participate in weaker intermolecular interactions. In the crystalline structure of cytosine, a molecule in the amino-oxo tautomeric form is linked to adjacent molecules by two strong hydrogen bonds $N(1')\text{--}H(1')\dots N(3)$ and two $N(4)\text{--}H(4)\dots O(2')$ bonds with lengths of 2.84, 2.98, and 3.03 Å, respectively [24]. These hydrogen-bonding interactions in cytosine are shorter and are therefore expected to be stronger than those found earlier in 2-TC, in agreement with the difference in melting points (593 versus 558 K).

The parameters that characterize the weak interactions in the crystalline structure of cytosine (shown in Fig. 7) are given in Table 4. Comparing the Laplacian isosurface and contours for cytosine (see Figs. 7 and 8) to those for 2-TC reveals the differences in VSCC localization. Four H-bonds—two type $NH\dots O$ and two type $CH\dots O$ —were found in cytosine, similar to the situation for 2-TC. The $N(4)\text{--}H(4)\dots O(2')$ bonds in cytosine have strengths of -14.08 and -14.13 kJ mol $^{-1}$, while those for $C(5)\text{--}H(5)\dots O(2')$ are -2.94 and -3.30 kJ mol $^{-1}$.

Although the two strongest hydrogen bonds involving oxygen from cytosine and sulfur from 2-TC are limited to a single octant in 3D space, the bonds with the oxygen from cytosine are weaker and hence the asymmetry parameter of the EFG tensor at ^{17}O is much greater (0.700 [62]) than that predicted by DFT at sulfur.

Although $-NH_2$ groups participate in three interactions with sulfur (oxygen in cytosine)—two hydrogen bonds with strengths of -17.12 and -10.62 kJ mol $^{-1}$ (-13.69 and

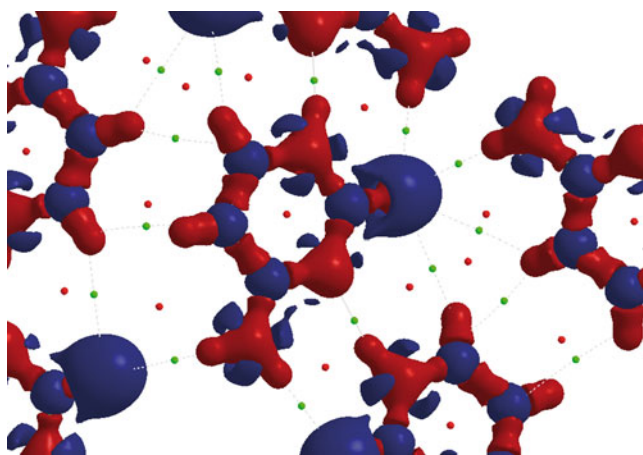


Fig. 7 3d distribution of the electron density Laplacian calculated by DFT for cytosine (isocontour ± 0.35 a.u.); the regions of negative Laplacian are shown in red and the regions of positive Laplacian in blue; small circles correspond to critical points (red RCP, green BCP)

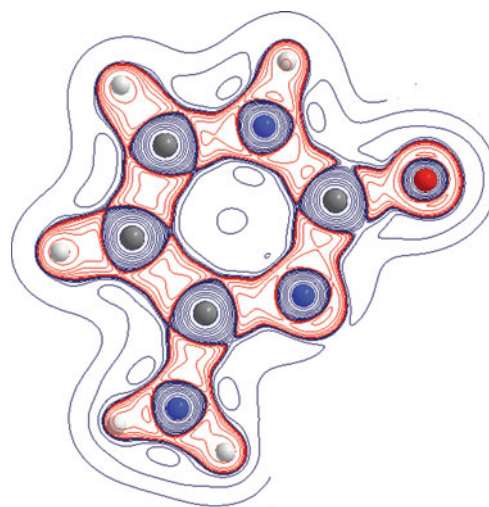


Fig. 8 Laplacian contours in cytosine (negative regions in red, positive regions in blue)

-13.53 kJ mol $^{-1}$) and an $S\dots N$ contact of strength -5.15 kJ mol $^{-1}$ (-0.53 kJ mol $^{-1}$), the hybridization and directions of the interactions are conserved, so the differences in the NQR parameters are relatively small. The $-N=$ and $-NH-$ sites participate in $NH\dots N$ bonds of strength -16.10 kJ mol $^{-1}$ (-35.12 kJ mol $^{-1}$ for cytosine), and the differences in the experimental η values for 2-TH and cytosine (see Table 1) lead to the conclusion that the asymmetry parameter η increases when the H-bond is shortened and hence also strengthened, which is in agreement with Seliger's [62] observation. It is worth noting that the quality of reproduction of the ^{14}N NQR parameters by DFT—assuming one of the most stable cytosine tautomers—is weaker than that for 2-TC, which is predominantly due to the poorer geometry but also the use of the older and thus less sophisticated method of collecting NQR data.

Although stacked $\pi\text{--}\pi$ interactions involving non-H atoms ($N\dots N$, $N\dots C$ and $N\dots O$ with lengths of 3.814, 3.392 and 3.668 Å and strengths of -1.85 , -3.50 and -3.35 kJ mol $^{-1}$)—which are purely van der Waals in nature—were detected in cytosine, they are weak. The stacking found in cytosine and that found in 2-TC involve generally similar overlapping, but the stacking $\pi\text{--}\pi$ interaction in cytosine is a bit stronger (-12.05 versus -9.7 kJ mol $^{-1}$), which is reflected mainly in the extra contact in cytosine ($N\dots C$) in comparison to 2-TC. Indeed, the improvement in the reproduction of the NQR parameters resulting from the $\pi\text{--}\pi$ stacking interaction is very important for cytosine (see Fig. 9) but practically negligible for 2-TC. Moreover, the size of the cluster has less of an effect on the quality of reproduction of the ^{17}O -NQR parameters (experimental data from [61]) than on that of the ^{14}N -NQR parameters (see Table 5, Fig. 9). It should be noted that the 2D or 3D cluster constructed in this manner does not

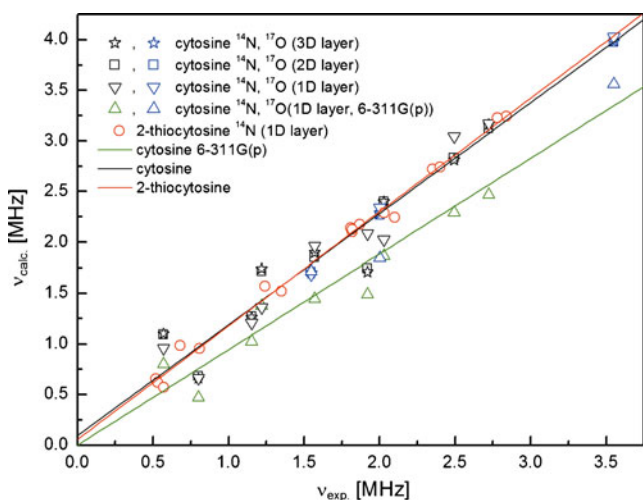


Fig. 9 The correlation between the experimental and calculated NQR frequencies for cytosine and 2-thiocytosine, showing the influence of π - π stacking

take into account periodic long-range effects. As follows from the QTAIM analysis, the differences between the electric charge distributions in the dimer and in the crystal are typically relatively small [60], so when a cluster built from 12 or 18 molecules is considered, the quality of reproduction of the electron density distribution is much improved. Moreover, according to the definition of the EFG tensor components, they decrease very quickly with increasing distance; e.g., towards the zz axis as x^2/r^5 . In view of the above, it can be assumed that long-range effects on the reproduction of the EFG tensor components are smaller than those that follow when stacking π - π interactions are accounted for (the latter does not exceed 5%), they do not affect the conclusions of this work.

The above analysis of intermolecular interactions permits one more conclusion to be drawn. As mentioned earlier, cytosine and 2-thiocytosine crystals differ in the number of nonequivalent molecules in an elementary cell (one versus two). The difference between these two structures (a difference that is ultimately due to replacing oxygen with sulfur) is responsible for the effects described above, including an elongation of the C-X (X=O, S) bond and the redistribution of the electron density. These effects cause the two compounds to exhibit slightly different intermolecular bonding patterns, as characterized above. However, it should be emphasized that the hierarchy of structures in crystals of 2-TC and cytosine molecules (supramolecular synthons (dimers), ribbons, stacks), which is reflected in progressively weaker bonds, is the result of the tendency to minimize polarization and enhance structural stability. Indeed, the values of the dipole moments of the monomer of cytosine and two nonequivalent monomers of 2-TC (7.31 versus 8.67 and 8.70 D) are much higher

Table 5 NQR parameters calculated at the B3LYP/6-311 G(d,p) and B3LYP/6-311 ++G(d,p) levels of theory for cytosine

Site	B3LYP/6-311 G(p)			B3LYP/6-311 ++G(d,p)			2D layer (12 molecules)			3D layer (18 molecules)				
	η	ν_+	ν_-	η	ν_+	ν_-	e^2Qq/h	ν_+	ν_-	ν_0	e^2Qq/h	ν_+	ν_-	ν_0
-NH ₂	2.520	0.635	2.290	1.490	0.800	1.490	3.045	2.085	0.959	3.047	0.717	2.831	1.739	1.092
-NH-	2.158	0.437	1.864	1.373	0.472	1.373	2.025	1.362	0.663	2.744	0.498	2.400	1.716	0.683
-N=	2.607	0.785	2.467	1.443	1.023	1.443	3.170	1.961	1.209	3.315	0.766	3.121	1.851	1.270
O	6.880	0.912	3.562	1.846	1.716	1.846	4.026	2.342	1.684	8.097	0.675	3.974	2.264	1.709

than those of the dimers (taking into account the most stable forms). However, the molecular synthon (dimer) of 2-thiocytosine consisting of two nonequivalent monomers is less polar than that of cytosine (4.83 versus 5.45D). This smaller polarity of the molecular synthon of 2-TC relative to that of cytosine is a consequence of the greater tendency of sulfur atoms than oxygen atoms to get involved in intermolecular interactions, which implies greater planarity of the molecular synthon of 2-TC than that of cytosine. Moreover, the arrangement of molecular synthons of 2-TC and cytosine in the neighboring ribbons guarantees a considerable—almost twofold—reduction in the dipole moment. A similar tendency to minimize polarization has been observed for the neighboring ribbons in 2-TC and cytosine. The 1D layer built from molecular synthons of 2-TC is slightly more polar than that made from molecular synthons of cytosine. Thus, the π - π stacking interaction between distinct layers in cytosine is slightly stronger than that in 2-TC. This greater steric repulsion for 2-TC leads to increased spacing between layers. It should be mentioned that, despite significant differences between their dipole moments, the differences between the orientations of their dipole moment vectors are small for monomers, dimers and 1D layers in cytosine and 2-TC. Thus, the ordering is generally similar in these solid-state structures, ignoring symmetry, which leads to a kind of duplication. It seems that the polar nature of the molecular synthons acts as a guiding template that produces a highly ordered solid-state structure in both cytosine and 2-thiocytosine.

Reactive sites in 2-thiocytosine and cytosine

In the molecules of 2-TC and cytosine, there are many sites that satisfy the criteria for reactive sites: they have one or more unshared pairs of electrons, polar bonds, and electron-deficient atoms or atoms with expandable octets. The Laplacian of the electron density can be treated as a function that probes these reactive sites. In Figs. 5 and 7, the regions in red in which the Laplacian is negative represent those in which the negative charge is concentrated, whereas the regions in blue are those in which the Laplacian is positive, characterized by a depletion of negative charge. The holes in the reactive surface defined as the isosurface, where the Laplacian changes its sign, clearly indicate that C(4) in 2-TC and C(2) and C(4) in cytosine are potential sites for nucleophilic attack. The electron-depleted regions at carbon atoms C(4)<C(2)<C(6) in 2-TC and C(2)<C(4)<C(6) in cytosine represent potential sites for nucleophilic attack. Additionally, in cytosine, the regions above and below C(2) are more vulnerable to potential nucleophilic attack than they are in 2-TC, where the claws over the nitrogen atom provide additional shielding.

The electron-rich regions at N(3)<N(4)<N(1)<S(2) in 2-TC and N(3)<N(4)<O(2)<N(1) in cytosine indicate potential sites for electrophilic attack.

It is worth noting that the value of the ^{14}N NQR frequency, which describes the degree of redistribution of the electron density from the nonbonding (lone) electron pair to equivalent bonds in which nitrogen participates, is highest for N(3), intermediate for N(2), and lowest for N(1), which clearly indicates the degree of susceptibility to electrophilic attack. Thus, the conclusions drawn from the theoretically calculated Laplacian are in a good agreement with the indications given by the NDQR results.

The amino-oxo tautomer of cytosine has a lower-lying HOMO (highest occupied molecular orbital) and a higher-lying LUMO (lowest unoccupied molecular orbital) than those of TC1 of 2-TC. The narrower HOMO–LUMO gap for the TC1 tautomer (3.56 eV) than for cytosine (5.09 eV) leads to a more stable structure for cytosine and a smaller excitation energy for TC1. In terms of the chemical reactivity descriptors (according to Parr's definition [63]), the chemical potential, absolute hardness and electrophilicity index are equal to -4.10 and -3.52 eV, 2.55 and 1.78 eV, and 3.30 and 3.47 eV for cytosine and TC1, respectively. The maximum electronic charge that cytosine can accept from the environment is equal to 0.81, while it is 0.99 for 2-TC. Consequently, cytosine—which is harder—should be less reactive than 2-TC in unimolecular reactions such as isomerization, dissociation and radical formation. This information could be helpful when attempting to rationally explain the activity of 2-TC in biological and chemical reactions. The enhanced reactivity of 2-TC compared to cytosine suggests the probable occurrence of point mutations in which cytosine is replaced by 2-thiocytosine, leading to reduced stability of DNA.

The HOMO–LUMO gap for the TC3 tautomer (3.69 eV) is only slightly wider than that for TC1 (3.56 eV), meaning that TC3 has a more kinetically stable structure than TC1. The chemical potential, absolute hardness and electrophilicity index are equal to -3.62 eV, 1.84 eV and 3.55 for TC3, while the maximum electronic charge that it can accept from the environment is 0.98. Thus, TC3—which is notably harder than TC1—should be less reactive than TC1 in unimolecular reactions. An analogous analysis performed for a cluster leads to completely different conclusions than those obtained for the isolated molecules. The cluster of TC1 tautomers of 2-TC has a lower-lying HOMO (highest occupied molecular orbital) and a higher-lying LUMO (lowest unoccupied molecular orbital) than those of the TC3 and cytosine clusters. The narrower the HOMO–LUMO gap in the TC3 tautomer (2.45 eV) is compared to that in the TC1 tautomer

(2.69 eV), the more stable the structure of TC1 and the smaller the excitation energy of TC3. In terms of the chemical reactivity descriptors, the chemical potential, absolute hardness and electrophilicity index are equal to -3.55 and -3.66 eV, 1.35 and 1.22 eV, and 4.68 and 5.49 eV for TC1 and TC3, respectively, while the maximum electronic charge that 2-TC can accept from the environment differs by 0.17 (1.32 and 1.49 e for TC1 and TC3, respectively). On the other hand, the slightly narrower HOMO–LUMO gap in cytosine (2.687 eV) compared to 2-TC means that the structure of 2-TC is more kinetically stable and the excitation energy of solid cytosine is smaller. The chemical reactivity descriptors—the chemical potential, absolute hardness and electrophilicity index—for cytosine are equal to -4.17 eV, 1.34 eV, and 6.48 eV, respectively, while the maximum electronic charge that cytosine can accept from the environment is 1.55 e. Intermolecular interactions in the solid cause the HOMO–LUMO gap for cytosine to be narrower than it is for 2-TC, which means that these interactions have a stronger stabilizing effect on the structure of cytosine than on that of 2-TC. This is also reflected in the slightly lower chemical hardness of cytosine (1.34 eV) than that of 2-TC (1.35 eV). These results also imply that in the solid phase, in contrast to the gas phase, sulfur is better than oxygen at stabilizing the structure by electron donation. Thus, solid-state properties have a profound impact on the stabilities and reactivities of both compounds.

Conclusions

1. Eighteen resonance lines detected by ^{14}N NQDR appear as a result of the nonequivalence of molecules in the elementary cell. According to the NQR spectra, in 2-TC one tautomeric form predominates, and the same is true of cytosine.
2. Scattering in the position of the proton—but not formal proton migration—is responsible for the “disordered” H-bonds detected in cytosine by FTIR and the relatively large scattering in the NQR parameters reported in the literature.
3. The combined substituent effect (replacement of oxygen with sulfur), which influences π -electron delocalization within the heterocyclic ring and changes the pattern of intermolecular interactions, is strongly manifested in the NQR spectra as a change in the number of lines in the spectrum (nonequivalent molecules) and changes in the values of NQR parameters of all ^{14}N atoms. These changes are relatively small at the $-\text{NH}_2$ site and significant at the $-\text{NH}-$ and $-\text{N}=\text{sites}$.

4. In cytosine (in contrast to 2-TC), the VSCC at the oxygen (sulfur) atom points towards the carbon atom, but there is practically no VSCC at the carbon that points towards the oxygen atom; the BCP at the $\text{C}=\text{O}$ bond is shifted towards the carbon basins (in contrast to 2-TC, in which it is located near the sulfur basins); and the electron densities at almost each BCP, excluding $\text{N}(3)-\text{C}(4)$ and $\text{C}(4)-\text{NH}_2$ in cytosine are higher than in 2-TC. This suggests that the π -electrons from $\text{C}=\text{O}$ to not only neighboring bonds but also the whole heterocyclic ring.
5. The orientation of the zz axis of the EFG tensor parallel to the bond path with the BCP at which the change of electron density is the highest explains the noticeable decrease in quadrupole coupling constant that occurs only at the $-\text{N}=\text{site}$ upon replacing sulfur with oxygen, while the more symmetrical distribution of the electron density in 2-TC compared to that in cytosine explains the increase in the NQR asymmetry parameter at the $-\text{NH}$ site.
6. The stacking $\pi-\pi$ interaction in cytosine is stronger, due to the larger spacing between layers in 2-TC compared with cytosine. Thus, the improvement in the reproduction of NQR parameters resulting from the $\pi-\pi$ stacking repulsive interaction is substantial for cytosine, but practically negligible for 2-TC.
7. The presence of a sulfur instead of an oxygen substituent at the $\text{C}(2)$ position is crucial, because the presence of this sulfur allows 2-TC to form much weaker hydrogen bonds that are reversible, which in turn facilitates biological processes (e.g., enzyme–substrate interactions) and can be used for regulation in metabolic processes. In the solid phase, in contrast to the gas phase, sulfur is better than oxygen at stabilizing the structure by electron donation.

Acknowledgments The generous allotment of computing time from the Poznan Supercomputing and Networking Center (PCSS) in Poznań, Poland is gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Carbon J, David H, Studier MH (1968) *Science* 161:1146–1147
2. Roosaf RA, DeLamater ED (1960) *Cancer Res* 20:1543–1554
3. Timson J, Price DJ (1972) In: Darlington CD, Lewis KR, Hafner PF (eds) *Chromosomes today*, vol 3. Wiley, New York, p 118
4. Timson J, Price DJ, Walker JS (1972) *Cytobios* 5:97–100
5. Lozzio CB, Wigler PW (1971) *J Cell Physiol* 78:25–31

6. Lozzio CB (1971) *Exp Cell Res* 69:377–383
7. Szinicz L, Albrecht GJ, Weger N (1981) *Arzneimittel Forschg* 31:1713–1717
8. Singh K, Groth-Vasselli B, Farnsworth PN, Rai DK (1996) *Res Commun Mol Pathol Pharmacol* 94:129–140
9. Podolyan Y, Gorb L, Blue A, Leszczynski J (2001) *J Mol Struct THEOCHEM* 549:101–109
10. Rostkowska H, Nowak MJ, Lapiński L, Bretner M, Kulikowski T, Les A, Adamowicz L (1993) *Biochim Biophys Acta* 1172:239–246
11. Kawaguchi T, Ichikawa T, Hasegawa T, Saneyoshi M, Yukita A, Asano M, Wakayama T, Kato H, Nagata T (1999) *Biol Pharm Bull* 22:100–102
12. Kawaguchi T, Ichikawa T, Hasegawa T, Saneyoshi M, Wakayama T, Kato H, Yukita A, Nagata T (2000) *Chem Pharm Bull* 48:454–457
13. Vetter C, Wagner Ch, Kaluderović GN, Paschke BR, Steinborn D (2009) *Inorg Chim Acta* 362:189–195
14. Beetz CP Jr, Ascarelli G (1980) *Spectrochim Acta A* 36:525–534
15. Yadav RA, Yadav PNS, Yadav JS (1988) *Spectrochim Acta A* 44:12011–1206
16. Shigeta S, Mori S, Kira T, Takahashi K, Kodama E, Konno K, Nagata T, Kato H, Wakayama T, Koike N, Saneyoshi M (1999) *Antivir Chem Chemother* 10:195–209
17. Bhave DP, Muse WB, Carroll KS (2007) *Infect Disord Drug Targets* 7:140–158
18. Nowak MJ, Lapiński L, Fulara J (1989) *Spectrochim Acta A* 45:229–242
19. Gould IR, Vincent MA, Hiller IH, Lapinski L, Nowak MJ (1992) *Spectrochim Acta A* 48:811–818
20. Jaworski A, Szczepaniak M, KiBulat K, Person WB (1990) *J Mol Struct* 223:63–92
21. Kobayashi R (1998) *J Phys Chem A* 102:10813–10817
22. Les A, Adamowicz L, Bartlett RJ (1989) *J Phys Chem* 93:4001–4007
23. Podolyan Y, Gorb L, Leszczynski J (2003) *Int J Mol Sci* 4:410–421
24. McClure RJ, Craven BM (1973) *Acta Crystallogr B* 29:1234–1238
25. Latosińska JN, Latosińska M, Koput J (2003) *J Mol Struct* 648:9–18
26. Rostkowska H, Nowak MJ, Lapiński L, Brenner M, Kulikowski T, Les A, Adamowicz L (1993) *Spectrochim Acta A* 49:551–565
27. Rostkowska H, Nowak MJ, Lapiński L, Brenner M, Kulikowski T, Les A, Adamowicz L (1993) *Biochim Biophys Acta* 1172:239
28. Kwiatkowski JS, Leszczynski J (1996) *J Phys Chem* 100:941–953
29. Kwiatkowski JS, Pullman B (1975) *Adv Heterocycl Chem* 18:199–335
30. Furberg S, Jensen LH (1970) *Acta Crystallogr B* 26:1260–1268
31. Rozenberg M, Shoham G, Reva I, Fausto R (2004) *Spectrochim Acta A* 60:463–470
32. Latosińska JN, Seliger J, Žagar V, Burchardt DV (2009) *J Phys Chem A* 113:8781–8790
33. Latosińska JN, Latosińska M, Seliger J, Žagar V, Kazimierczuk Z (2009) *Chem Phys Lett* 476:293–302
34. Latosińska JN, Latosińska M, Seliger J, Žagar V, Maurin J, Orzeszko A, Kazimierczuk Z (2010) *J Phys Chem A* 114:563–575
35. Hansen NK, Coppens P (1978) *Acta Crystallogr A* 34:909–921
36. Allen FH (1986) *Acta Crystallogr B* 42:515–522
37. Hammond GS (1955) *J Am Chem Soc* 77:334–338
38. Seliger J (2000) In: Lindon JC, Tranter GE, Holmes JL (eds) *NQR theory in encyclopedia of spectroscopy and spectrometry*. Academic, San Diego, pp 1672–1680
39. Seliger J, Osredkar R, Mali M, Blinc R (1976) *J Chem Phys* 65:2887–2893
40. Seliger J, Blinc R, Arend H, Kind R (1976) *Z Phys B* 25:185–189
41. Stephenson D, Smith JAS (1988) *Proc R Soc Lond A* 416:149–178
42. Seliger J, Žagar V (2008) *J Magn Reson* 193:54–62
43. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven T Jr, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, M. A. Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) *Gaussian 03*, revision D.01. Gaussian Inc., Wallingford
44. Becke AD (1993) *J Chem Phys* 98:1372–1377
45. Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37:785–789
46. Latosińska JN (2000) *Magn Reson Chem* 38:192–196
47. Bader RFW (1994) *Atoms in molecules: a quantum theory*. Oxford University Press, Oxford
48. Espinosa E, Molins E, Lecomte C (1998) *Chem Phys Lett* 285:170–173
49. Garcia MLS, Smith JAS (1983) *J Chem Soc Perkin Trans II* 1401–1408
50. Blinc R, Mali M, Osredkar R, Prelesnik A, Seliger J, Zupancic I, Ehrenberg L (1972) *J Chem Phys* 57:5087–5093
51. Saedinianiach azota (Nuclear quadrupole resonance in nitrogen compounds). Nauka, Moscow
52. Klein O, Aguilar-Parrilla F, del Lopez Amo JM, Jagerovic N, Elguero J, Limbach HH (2004) *J Am Chem Soc* 126:11718–11732
53. Desiraju GR (1995) *Angew Chem Int* 34:2311–2327
54. Koch PLA, Popelier A (1995) *J Phys Chem* 99:9747–9754
55. Gatti C (2005) *Z Kristallogr* 220:399–457
56. Bader RFW, Esse H (1984) *J Chem Phys* 80:1943–1960
57. Espinosa E, Alkorta I, Elguero J, Molins E (2002) *J Chem Phys* 117:5529–5542
58. Macchi P, Proserpio DM, Sironi A (1998) *J Am Chem Soc* 120:13429–13435
59. Jeffrey GA (1997) *An introduction to hydrogen bonding*. In: Truhlar DG (ed) *Topics in physical chemistry*. Oxford University Press, New York, p 303
60. Wu G, Dong S, Ida R, Reen N (2002) *J Am Chem Soc* 124:1768–1777
61. Munshi P, Cameron E, Cameron TS, Row TNG, Ferrara JD (2007) *J Phys Chem A* 111:7888–7897
62. Seliger J (1998) *Chem Phys* 231:81–86
63. Parr RG, Pearson RG (1983) *J Am Chem Soc* 105:7512–7516

Molecular modeling, docking and dynamics simulations of GNA-related lectins for potential prevention of influenza virus (H1N1)

Huai-long Xu · Chun-yang Li · Xue-mei He ·
Ke-qin Niu · Hao Peng · Wen-wen Li ·
Cheng-cheng Zhou · Jin-ku Bao

Received: 21 October 2010 / Accepted: 13 February 2011 / Published online: 29 March 2011
© Springer-Verlag 2011

Abstract The *Galanthus nivalis* agglutinin (GNA)-related lectin family exhibit significant anti-HIV and anti-HSV properties that are closely related to their carbohydrate-binding activities. However, there is still no conclusive evidence that GNA-related lectins possess anti-influenza properties. The hemagglutinin (HA) of influenza virus is a surface protein that is involved in binding host cell sialic acid during the early stages of infection. Herein, we studied the 3D-QSARs (three-dimensional quantitative structure–activity relationships) of lectin– and HA–sialic acid by molecular modeling. The affinities and stabilities of lectin– and HA–sialic acid complexes were also assessed by molecular docking and molecular dynamics simulations. Finally, anti-influenza GNA-related lectins that possess stable conformations and higher binding affinities for sialic acid than HAs of human influenza virus were screened, and a possible mechanism was proposed. Accordingly, our results indicate that some GNA-related lectins, such as *Yucca filamentosa* lectin and *Polygonatum cyrtonema* lectin, could act as drugs that prevent influenza virus infection via competitive binding. In conclusion, the GNA-related lectin family may be helpful in the design of novel candidate agents for preventing influenza A infection through the use of competitive combination against sialic acid specific viral infection.

Keywords *Galanthus nivalis* agglutinin (GNA)-related lectins · Hemagglutinin (HA) · Influenza A virus · *Polygonatum cyrtonema* lectin (PCL) · Viral infection

Abbreviations

GNA	<i>Galanthus nivalis</i> agglutinin
TGL	<i>Tulipa gesneriana</i> lectin
YFL-II	<i>Yucca filamentosa</i> lectin
YFL-I	<i>Yucca filamentosa</i> lectin
3D-QSAR	Three-dimensional quantitative structure–activity relationship
AML ^a	<i>Arum maculatum</i> lectin
AAL	<i>Arisaema amurense</i> lectin
PLC	<i>Pinellia cordata</i> lectin
AML ^b	<i>Alocasia macrorrhiza</i> lectin
PTL	<i>Pinellia ternata</i> lectin
PCL	<i>Polygonatum cyrtonema</i> lectin
AHL	<i>Arisaema heterophyllum</i> lectin
HA	Hemagglutinin
HA-I	1934 Human H1 HA
HA-II	1918 Human H1 HA
vRNP	Viral nucleoprotein

Introduction

Plant lectins are an important type of glycoprotein that are widespread throughout the plant kingdom and possess the ability to recognize carbohydrate moieties specifically as well as to bind with carbohydrates reversibly [1–3]. The *Galanthus nivalis* agglutinin (GNA)-related lectin family, a significant family of plant lectins, has drawn increasing attention in recent years due to its remarkable antitumor and antiviral properties. These notable biological properties of the GNA-

Huai-long Xu and Chun-yang Li contributed equally to this work

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1022-7) contains supplementary material, which is available to authorized users.

H.-l. Xu · C.-y. Li · X.-m. He · K.-q. Niu · H. Peng · W.-w. Li ·
C.-c. Zhou · J.-k. Bao (✉)
School of Life Sciences, Sichuan University,
Chengdu 610064, China
e-mail: jinkubao@yahoo.com

related lectin family are closely related to their carbohydrate-binding activities [4–6]. Most GNA-related lectins exhibit mannose-binding activity, while several lectins such as PCL (*Polygonatum cyrtonema* lectin) [7] and OJL (*Ophiopogon japonicus* lectin) [8] also display carbohydrate affinity towards sialic acid. Several GNA-related lectins exhibit significant antiviral properties, such as the anti-HIV activity of GNA [4] and PCL [9] as well as the anti-HSV-II effect of OJL [8, 10]. Whether GNA-related lectins can also show similar anti-influenza properties remains to be verified.

The surface of the influenza A virus is made up of diverse proteins; among these, the most important envelope glycoprotein implicated in receptor recognition and fusion is hemagglutinin (HA). Another essential glycoprotein is neuraminidase (NA), which assists HA in binding with the corresponding receptor and initiating infection [11]. HA exhibits a variety of distinct structures, and all of these different types of HA can bind receptors containing glycans with terminal sialic acids [11–13]. The combination of HA with sialylated glycans exposed on the host cell surface is the key first step in the infection, transmission and virulence of influenza viruses [14, 15]. Although HA must accumulate mutations rapidly and continuously to escape immune system recognition, one recent report verified that HA of the 2009 influenza A virus was extremely similar to the HA of the 1918 influenza A virus [16]. As the influenza A virus has mutated, it has infected species ranging from avians to humans; avians are mainly infected by α 2,3-linked sialic acid specific influenza A virus strains while humans are mainly infected by α 2,6-linked sialic acid specific influenza A virus strains [17–21].

In the current study, we modeled the three-dimensional structures of GNA-related lectins and assessed the affinities and stabilities of lectin– and HA–sialic acid complexes via molecular docking and dynamics simulations. In addition, we screened potential GNA-related lectins against human influenza virus, and proposed possible mechanisms for the activity of such lectins toward human influenza virus. Accordingly, our results indicate that some GNA-related lectins would be promising candidate agents for preventing influenza virus infections, as they would competitively bind with receptors containing glycans with terminal sialic acids. Thus, these findings may lead to the exploration of the GNA-related lectin family as potential agents against various sialic acid specific viral infections in the future.

Materials and methods

Data preparation

The structures of two representative HAs were retrieved from the Protein Data Bank (PDB) (<http://www.pdb.org/pdb/home/>

[home.do](http://www.pdb.org/pdb/home/)); the PDB codes were 1RU7 (1934 Human H1 Hemagglutinin) and 1RUZ (1918 H1 Hemagglutinin), respectively. The sequences of GNA-related lectins were acquired from Universal Protein Resource (<http://www.uniprot.org>) and one study performed by Els [22]. All of the sequences were further screened, and finally 51 potential sialic acid binding GNA-related lectins were obtained. The sialic acid structure was separated from HA of the 1934 influenza A virus in complex with a known human receptor analog, LSTc (PDB ID: 1RVZ), using the Chimera program [23].

Molecular modeling and structural alignment

Domains of these lectin sequences were acquired by Prosite (<http://www.expasy.ch/prosite/>) [24], and these domains were further utilized to construct the three-dimensional structures of 51 GNA-related lectins using MODELLER9v7 software [25], with the structure of GNA (*Galanthus nivalis*) lectin, PDB ID: 1JPC [26], used as the template. All structures were optimized using the Chimera program; optimization included hydrogen atom additions and energy minimizations, as well as the elimination of bound crystal water molecules or other organic compounds [23]. The Chimera MatchMaker tool [23] was utilized to compare the homology modeling structure and the crystal structure of PCL retrieved from the Protein Data Bank (PDB ID: 3A0C [39]).

Docking experiment

All molecular docking calculations were performed using the UCSF DOCK6.3 program [27] with a flexible ligand docking to a rigid receptor in a grid-based scoring method. The flexible ligand docking algorithm in DOCK6.3 allows the ligand (sialic acid) to structurally rearrange in response to the receptor (GNA-related lectin and HA) using the anchor-and-grow feature, which first identifies the largest rigid substructure (the anchor) and the flexible layers of the ligand, and then builds the flexible layers of the ligand onto the best anchor orientations within the context of the receptor (the growth stage). In our docking process, we increased the maximum number of orientations to 1000.

Molecular dynamics (MD) simulations

Molecular dynamics simulations were performed using the GROMACS (version 4.0.5) software package [28], with the GROMOS96 43a2 force-field [29] and a single-point-charge (spc) water model [30]. The sialic acid topology was generated using the Dundee PRODRG2 server [31], and the simulations of HA– and lectin–sialic acid complexes were performed under the same conditions.

Long-range electrostatic interactions were computed by a particle mesh Ewald (PME) algorithm at every step [32,

33]. The linear constraint (LINCS) [34] and Berendsen coupling [35] algorithms were utilized to constrain the lengths of all bonds and to keep the pressure (1 bar) and temperature (300 K) constant, so that a time step of 2 fs could be employed under isothermal–isobaric conditions. The cutoff distance used for van der Waals (VdW) forces was 1.4 nm, and the coordinates were saved every 500 steps. The resulting trajectory files were viewed and analyzed using VMD software [36].

Results

Overall structures of GNA-related lectins and hemagglutinin

Most GNA-related lectins are mannose specific, and the significant conserved “QXDXNXVXY” motif of the subdomain plays a crucial role in mannose recognition [8, 37, 38]. However, several lectins also exhibit sialic acid binding activity that may result from an amino acid mutation in the conservative mannose-binding motif of GNA-related lectins. Therefore, all of the sequences were further screened, and 51 GNA-related lectins (as shown in Fig. S1 of the “Electronic supplementary material,” ESM) possessing potential sialic acid specificity were finally obtained.

It has been reported that the monomers of GNA-related lectins have similar three-dimensional structures containing three subdomains. Each subdomain is made up of three or four strands of antiparallel β -sheets connected by Ω -loops. All of the 51 lectins exhibited similar three-dimensional structures (data not shown), and possessed sialic acid binding activity. For example, in YFL-I (*Yucca filamentosa* lectin) and AML^a (*Arum maculatum* lectin), the three-dimensional structures of YFL-I and AML^a (see Fig. 1a and c) display similar conformations. The motions of the YFL-I–Sia complex are also shown in Video S1 of the ESM.

To further validate the abovementioned results, we retrieved the crystal structure (determined by X-ray diffraction; PDB code: 3A0C) of the PCL with anti-HIV and mannose-binding activities [39]. Figure 2 clearly shows that the PCL structure obtained by homology modeling bears a close resemblance to the three-dimensional structure of PCL obtained by X-ray diffraction, with a RMSD in the C α position of 0.680 Å. This comparison indicates that the structure from homology modeling is very similar to the native crystal structure, which supports the hypothesis mentioned above.

In addition, sialic acid binding sites of receptor HA consisted mainly of the 190 helix (residues 190 to 198), the 130 loop (residues 135 to 138) and the 220 loop (residues 221–228). Our docking results illustrated that sialic acid could bind to these sites, with the HA-I (1934 Human H1 HA)–Sia complex forming four hydrogen bonds involving His180,

Glu212, Asn227 and Tyr229 (Table 1 and Fig. 1e). The HA-II (1918 Human H1 HA) system formed three hydrogen bonds involving Glu213, Arg217 and Arg226 (Table 1 and Fig. 1f). HA-I in complex with sialic acid is visualized in the video S2 in the ESM.

Binding energies

Accordingly, we docked two HAs of human influenza virus [PDB codes: 1RU7 (HA-I) and 1RUZ (HA-II)], as well as 51 GNA-related lectins with sialic acid. Ten GNA-related lectins showed greater affinities for sialic acid than the HAs, as shown in Table 1 (these structures are shown in Fig. 1 and Fig. S2 of the ESM). The binding energies of these ten GNA-related lectins in complex with sialic acid range from -41.15 kcal mol⁻¹ to -46.55 kcal mol⁻¹, although *Tulipa gesneriana* lectin (TGL) possesses the highest specificity for sialic acid. The binding energies of the two HA–Sia complexes were -40.84 kcal mol⁻¹ and -40.85 kcal mol⁻¹, respectively. Consistent with the molecular dynamics simulation results, the average total energies of the lectin– and HA–Sia complexes also indicated that the specificities of the two HAs for sialic acid were similar, and that TGL possessed the highest affinity for sialic acid (as shown in Table 1).

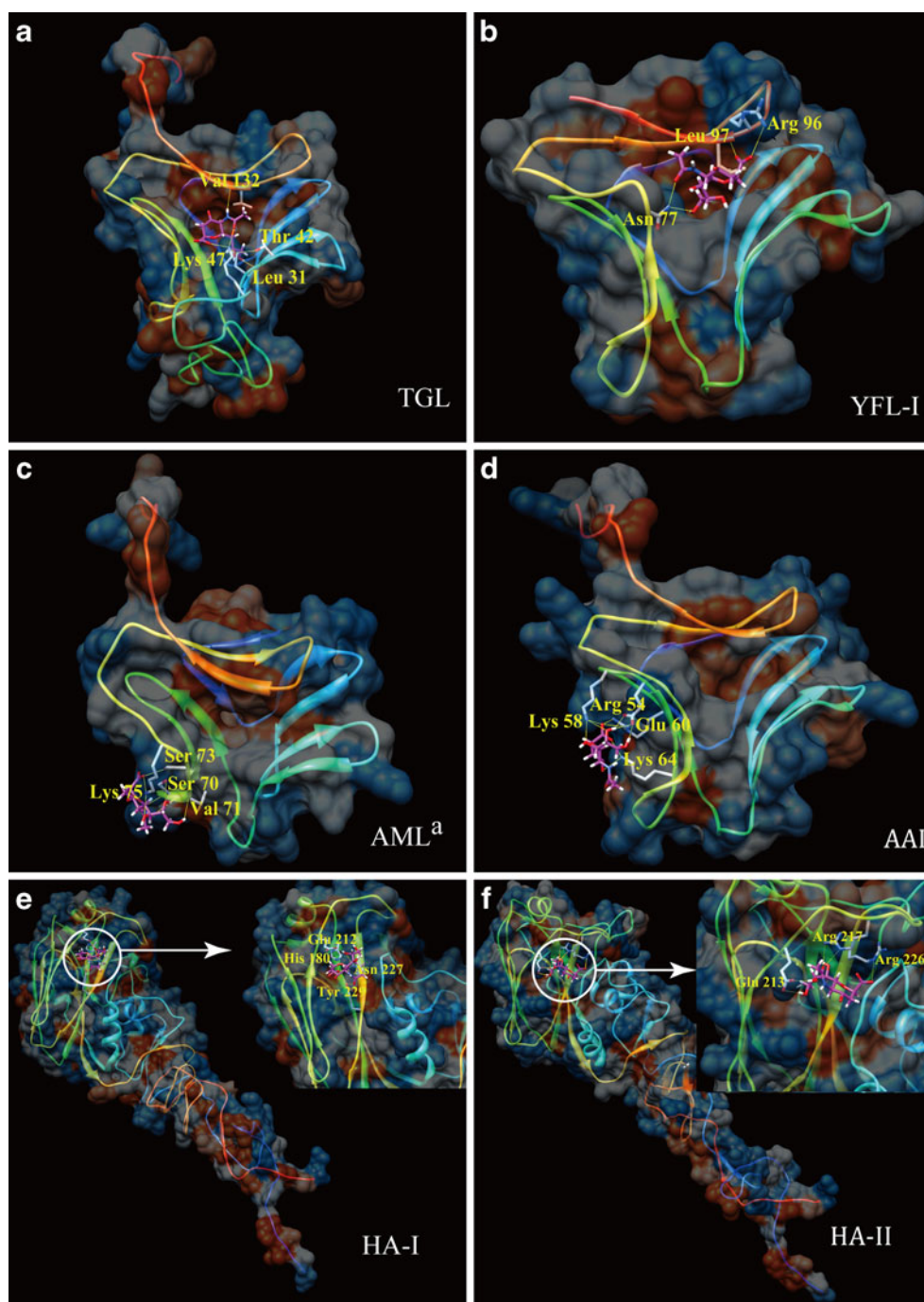
Further investigations of the overall structures of GNA-related lectin and sialic acid conjugates demonstrated that there were two distinct binding types of sialic acid, which meant that the GNA-related lectins could be largely divided into two groups. For instance, TGL– and YFL–Sia belonged to the first binding type, where the sialic acid binds to the center of the lectin (Fig. 1a, b and Fig. S2 of the ESM). Other complexes such as AML^a–Sia, *Arisaema amurense* lectin (AAL)–Sia, and *Polygonatum cyrtoneuma* lectin (PCL)–Sia belong to the second binding type, where sialic acid is almost bound to the second subdomain of the GNA-related lectin (Fig. 1c, d and Fig. S2 of the ESM).

Exploration of the binding sites of the GNA-related lectin–Sia complexes showed that the second conserved motif of “QXDXNXVXY” was mainly implicated in sialic acid recognition; in this motif, the majority of the key residues were K, E and K, which had mutated from the conserved D, N and Y residues, respectively. The R or T residues of several GNA-related lectins may also be involved in sialic acid binding (Table 1 and Fig. 3).

Root mean square deviations

To evaluate the stabilities of lectin– and HA–Sia complexes during the MD simulations, we calculated the root mean square deviation (RMSD) with respect to the initial structures along the 1.2 ns trajectory (shown in Fig. 4). These results indicated that the first binding type of lectin–

Fig. 1 The overall modeling of protein (GNA-related lectins and HAs)–sialic acid complexes: TGL–sialic acid (**a**), YFL-I–sialic acid (**b**), AML^a–sialic acid (**c**), AAL–sialic acid (**d**), 1934 Human H1 HA–sialic acid (**e**), and 1918 Human H1 HA–sialic acid (**f**). All complexes are presented in surface potential mode and stickball mode. In stickball mode, the sialic acid is represented by a *purple stick*, while the hydrogen bonds are shown as *yellow lines*



Sia complexes could reach equilibration and oscillate around an average value. YFL-I possessed the lowest RMSD and reached the plateau value after about 200 ps of simulation time. YFL-II reached its plateau value at a similar time, but the conformation of the YFL-II–Sia system was found to be less stable than that of YFL-I, whereas the TGL–Sia system showed more variability.

In addition, RMSD analyses of the second binding type of lectin–Sia complexes illustrated that three [AHL (*Arisaema*

heterophyllum lectin), AML^b (*Alocasia macrorrhiza* lectin) and PCL] of the seven GNA-related lectins equilibrated quickly after about 550 ps of simulation time, while the other four complexes were more variable. AHL was able to reach its plateau value of about 0.45 nm after 250 ps simulation time, and it maintained a stable conformation. Figure 4c indicates that AHL–Sia is more stable than the corresponding complexes of the other six lectins, despite the fact that its RMSD value was not the lowest. However,

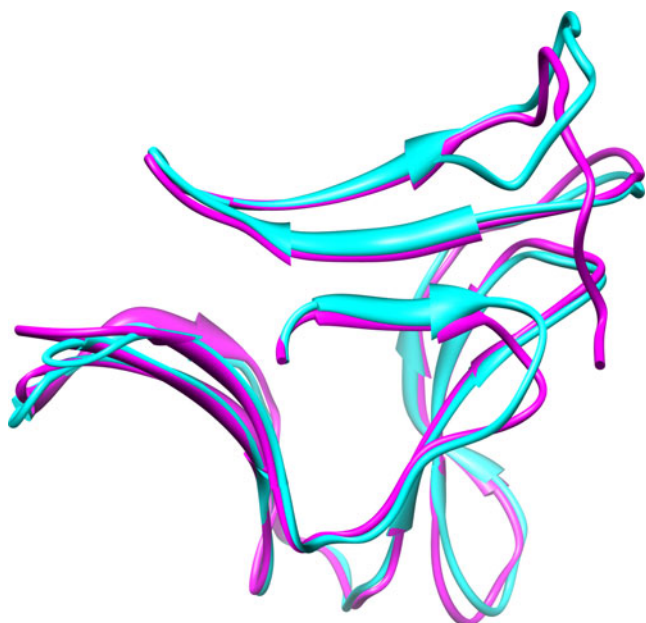


Fig. 2 Structural similarity between the structure obtained from homology modeling and the crystal structure of PCL. The structure of PCL from homology modeling is colored *cyan*, while its crystal structure is colored *magenta*

the RMSD results for HA–sialic acid complexes demonstrated that they were more variable during the MD simulations. These results indicate that lectin–Sia complexes are more stable than HA–Sia complexes.

The secondary structure

To further gauge the stabilities of lectin– and HA–Sia complexes, the variation in the secondary structure was

analyzed across 1.2 ns trajectories (as shown in Fig. 5 and Fig. S3 of the ESM). The results clearly show that the conformations of lectin– and HA–Sia remained stable during the 1.2 ns MD simulations. As mentioned above, GNA-related lectins are made up of three or four antiparallel β -sheets connected by Ω -loops. Secondary structural analyses also showed that there were several slight variations in the lectin–Sia conformations during MD simulations. For instance, residues 78–82 of PCL (the β -sheet) changed from a coil to a β -sheet during the MD simulations. However, lectin–Sia binding domains, such as residues 96–97 of YFL-I and residues 70–75 of AML^a, were relatively stable during MD simulation.

Although HA–Sia complexes and binding domains between HAs and sialic acid also remained relatively stable during the 1.2 ns MD simulations, several slight variations disturbed the stabilities of these two systems, such as those associated with residues 210–229 of HA-I (shown in Fig. 5 and Fig. S3 of the ESM).

Hydrogen-bonding analysis

As it is one of the most important forces that maintains the binding between ligand and receptor, hydrogen bonding can also reflect complex stability. Thus, we analyzed the changes in the hydrogen bonding in the the binding domain between proteins (HAs or lectins) and sialic acid (shown in Fig. 6 and Fig. S4 of the ESM).

HA-I in complex with sialic acid reached equilibrium at 480–1200 ps, and the average number of hydrogen bonds between HA-I and sialic acid was 1.52. Although complexes of lectins such as AML^a and PLC (*Pinellia cordata* lectin) with sialic acid presented more hydrogen bonds than

Table 1 Results of molecular docking and molecular dynamics simulations of GNA-related lectins and HA in complex with sialic acid

Protein name	Organism	Binding energy ^a	Total energy ^b	Hydrogen bonds ^a	Binding type
TGL	<i>Tulipa gesneriana</i>	−46.55	−363977	Lys47, Val132, Thr42, Leu31	I
YFL-II	<i>Yucca filamentosa</i>	−44.10	−162093	Lys81, Leu103	I
YFL-I	<i>Yucca filamentosa</i>	−43.92	−225707	Asn77, Arg96, Leu97	I
AML ^a	<i>Arum maculatum</i> (Cuckoo-pint)	−43.95	−285397	Lys75, Ser73, Ser70, Val71	II
AAL	<i>Arisaema amurense</i>	−43.90	−327434	Glu60, Lys58, Arg54, Lys64	II
PLC	<i>Pinellia cordata</i>	−42.52	−279799	Asp55, Thr54, Thr52, Arg56, Lys62	II
AML ^b	<i>Alocasia macrorrhiza</i> (Giant taro)	−41.86	−228785	Lys64, Asn56, Lys58, Glu60	II
PTL	<i>Pinellia ternata</i>	−41.42	−264601	Glu60, Lys64	II
PCL	<i>Polygonatum cyrtonema</i>	−41.27	−187606	Arg54, Thr56, His58, Asn59	II
AHL	<i>Arisaema heterophyllum</i>	−41.15	−239089	Glu58, Lys62, Arg66	II
HA-I	H1N1	−40.84	−2.49×10 ⁶	His 180, Glu 212, Asn 227, Tyr 229	–
HA-II	H1N1	−40.85	−2.72×10 ⁶	Glu 213, Arg 217, Arg 226	–

^a indicates molecular docking results

^b indicates molecular dynamics simulation results

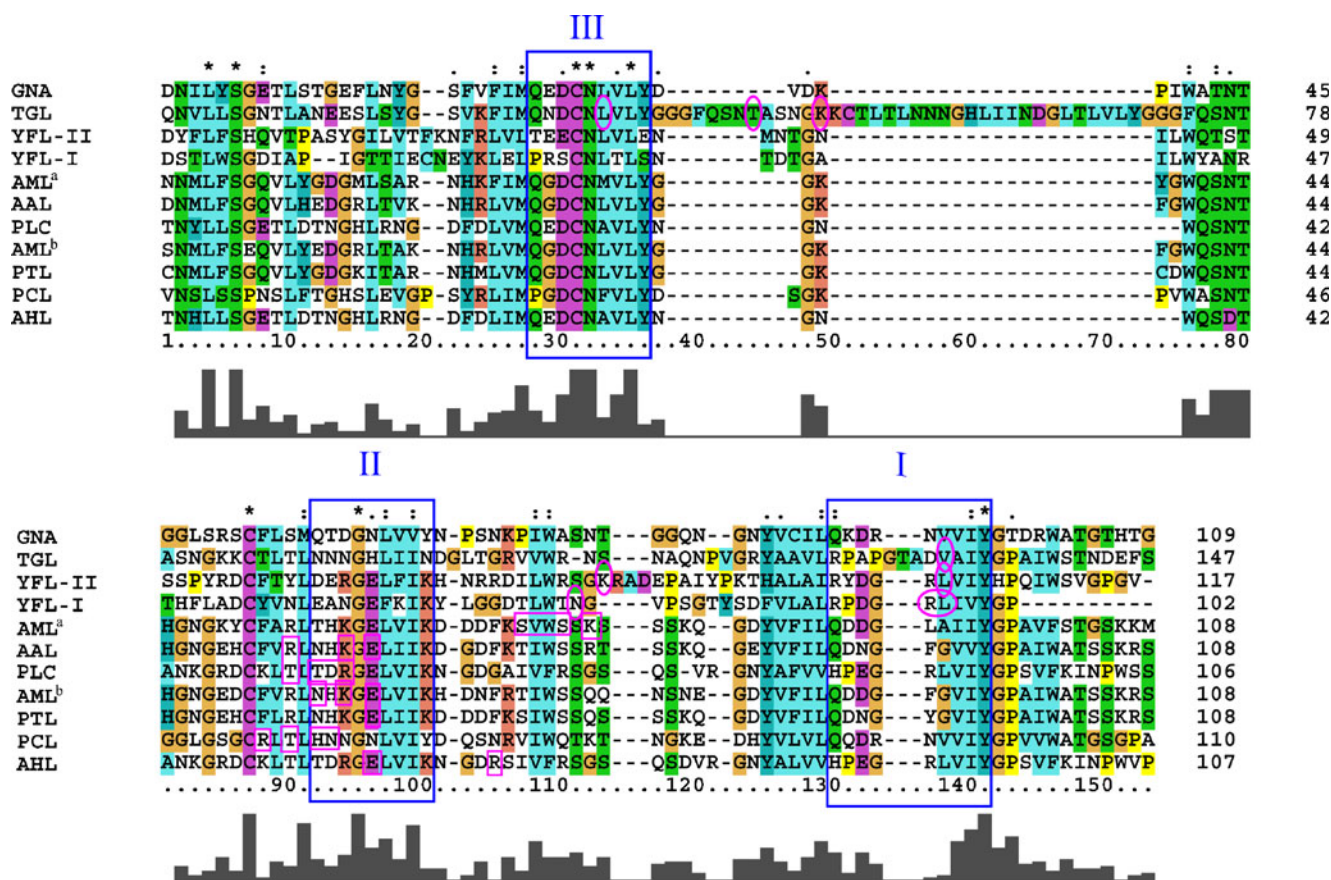


Fig. 3 Sequence alignment of GNA with ten GNA-related lectins that exhibit significantly greater affinity for sialic acid than HA. GNA-related lectins show significant sequence similarities, and the second conserved motif is mainly implicated in sialic acid recognition

HA complexed with sialic acid (as shown in Fig. 6d), the conformations of AML^a and PLC in complex with sialic acid tended to change more, according to RMSD analyses.

The RMSD analyses also showed that the average number of hydrogen bonds in the YFL-I–Sia complex was similar to the number of H-bonds in the HA–Sia system (as shown in Fig. 6b, d), but YFL-I exhibited a more stable structure. All of these results suggest that YFL-I may show more affinity than HA towards sialic acid.

Discussion

The GNA-related lectin family shows structural conservation and is mannose specific. The significantly conserved “QXDXNXVXY” motif of the subdomain plays a crucial role in mannose recognition [8, 37, 38]. In addition to their affinity to mannose, several GNA-related lectins can also bind with sialic acid, which may be due to the mutation of the conservative mannose-binding motif [9]. These affinities for mannose and sialic acid are also closely related to the antitumor and antiviral properties of GNA-related lectins [4, 8].

The surface glycoprotein hemagglutinin (HA) of influenza A virus is specific to sialic acid, so it may infect the host by binding with receptors that contain glycans with terminal sialic acids [11–13]. Therefore, it is urgently necessary to study the competitive combination of GNA-related lectins and sialic acid with view to its potential antivirus activity, which mainly occurs through the mechanism of steric hindrance. The experiments described in this paper were based on a combined approach involving molecular modeling, docking and molecular dynamics screening of potential GNA-related lectins [40–42] that possess stable conformations when in complex with sialic acid, and higher affinities than HAs for sialic acid.

In our study, ten GNA-related lectins were found to have higher affinities than HAs for sialic acid, and the TGL–Sia and YFL-I–sialic acid complexes exhibited relatively strong binding activities according to molecular docking results. Consistent with these docking results, the average total complex energy—which is a good indicator of the stability of the system—suggested that TGL and YFL-I show the greatest specificities for sialic acid.

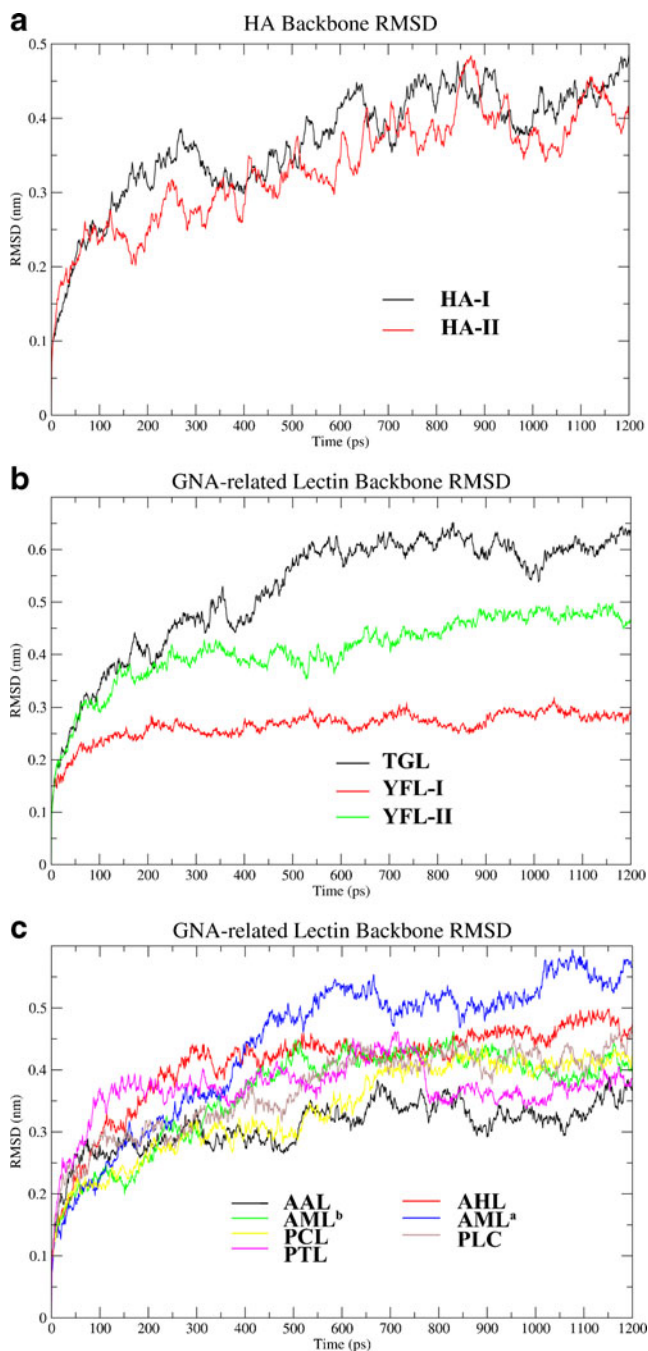


Fig. 4 Root mean square deviations of protein (GNA-related lectins and HAs)-sialic acid complexes over time from molecular dynamics simulations for HA-sialic acid complexes (a), the first binding type of GNA-related lectin-sialic acid complexes (b), and the second binding type of GNA-related lectin-sialic acid complexes (c)

Moreover, we assessed the stability of each of the ten systems by analyzing the root mean square deviation (RMSD), which indicates the Euclidean distance from the structure to a reference structure like the crystal structure or backbone, and only energy-stable complexes were accepted. RMSD results showed that most of the lectin-Sia complexes maintained a stable conformation throughout the 1.2 ns

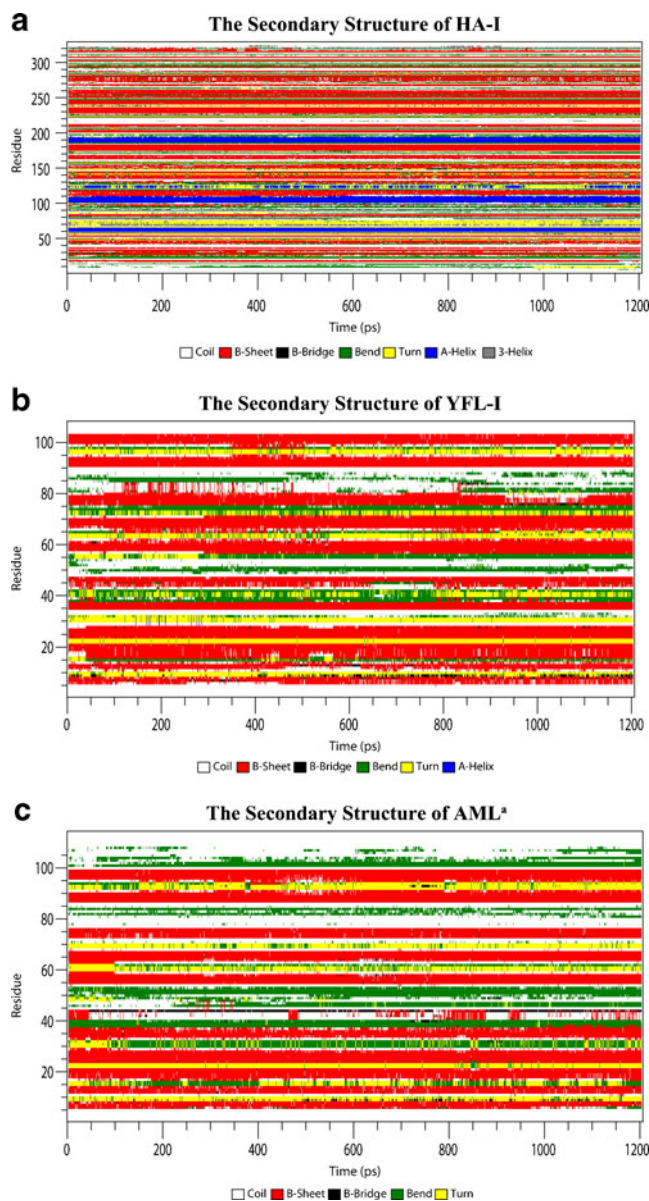


Fig. 5 Secondary structural variations of protein (GNA-related lectins and HAs)-sialic acid complexes in molecular dynamics simulations. Secondary structural variations over time for: 1934 Human H1 HA-sialic acid (a), YFL-I-sialic acid (b) and AML^a-sialic acid (c) complexes

trajectory, while the YFL-I system exhibited the greatest stability. Besides the RMSD, secondary structural and hydrogen-bonding analyses can also depict the stability of such a system. In accord with the RMSD analyses, most of the lectin-sialic acid complexes exhibited relatively stable conformations throughout the simulations, although the YFL-I complex possessed the greatest stability.

If the lectin-Sia structure is stable and presents a greater affinity than HA for sialic acid, this GNA-related lectin could be screened for potential anti-H1N1 activity due to competitive binding. YFL-I-Sia complex seemed to be the least stable structure according to stability analyses, but

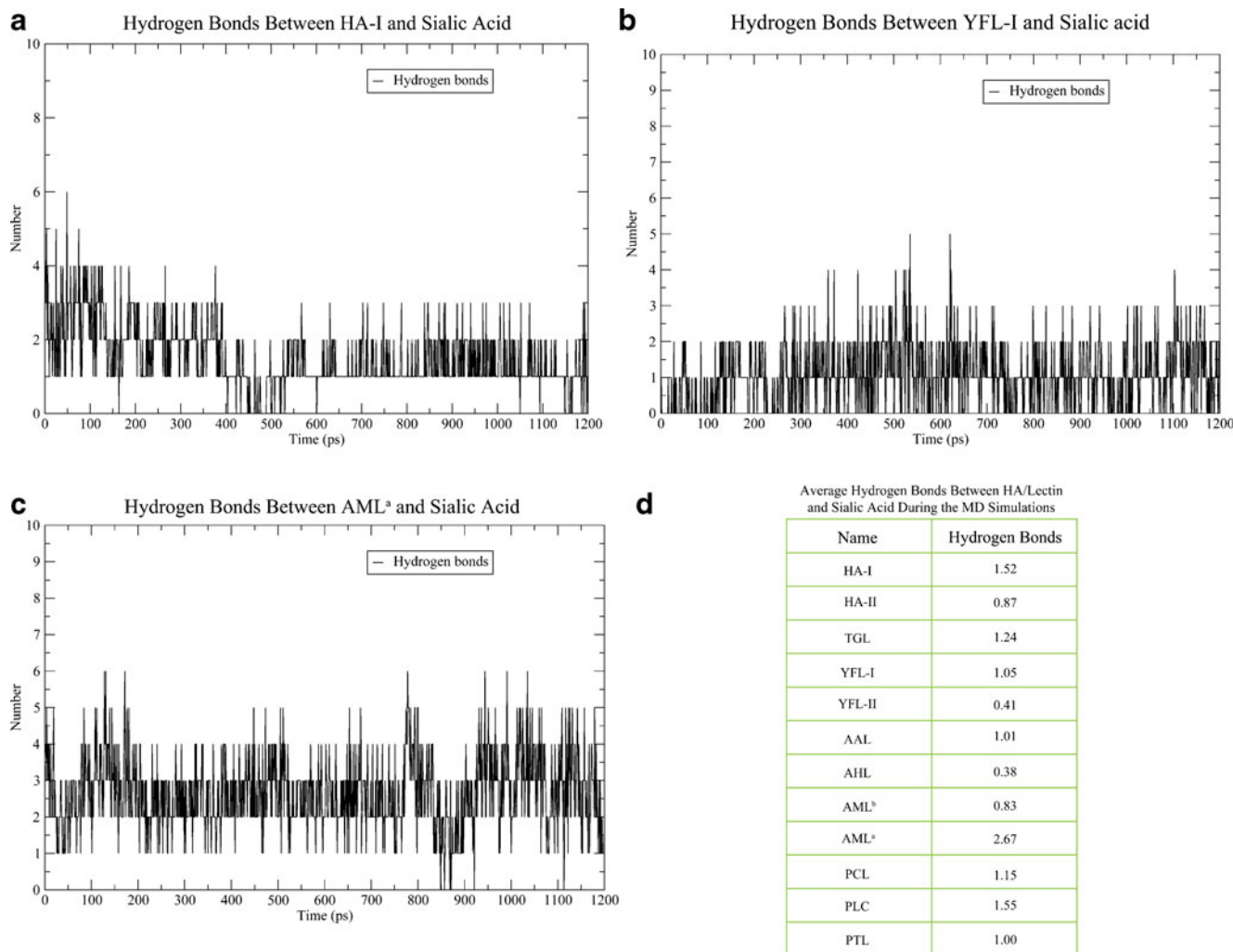


Fig. 6 Hydrogen-bonding variations in protein (GNA-related lectins and HAs)–sialic acid complexes during molecular dynamics simulations: 1934 Human H1 HA–sialic acid (**a**), YFL-I–sialic acid (**b**) and AML^a–

sialic acid (**c**) complexes. **d** Average number of hydrogen bonds between HAs/lectins and sialic acid during the course of the simulation

YFL-I possessed a relatively strong affinity for sialic acid. TGL in complex with sialic acid presents the strongest specificity, while this system is not the most stable. These differences between the complexes may arise from different binding domains and overall structures. The reason that the TGL–Sia complex had the strongest affinity but not the most stable structure may be the long sequence length of TGL, which might lead to the generation of novel and different overall structures. The complex with AML^a possessed more hydrogen bonds than that with YFL-I, which may be due to amino acid mutations in the carbohydrate recognition motif. In addition, structural uncertainties and inaccuracies of the docking and MD simulations could lead to these different results.

As mentioned above, YFL-I was found to have the greatest potential for use as a novel anti-influenza drug owing to the stable YFL-I–Sia conformation and the relatively strong

affinity of YFL-I in complex with sialic acid. Accordingly, members of the GNA-related lectin family, which show similar properties to YFL-I, may be candidate drugs to fight influenza.

Recent reports have demonstrated that several GNA-related lectins possess notable antiviral activities; for example, the anti-HIV activities of GNA [4] and PCL [9], as well as the anti-HSV-II effect of OJL [8, 10]. GNA and PCL exert their anti-HIV activities by blocking the binding between HIV envelope glycoprotein gp120 and the corresponding receptor CD4 [4, 43]. The glycans of gp120 consist of a high-mannose type and a complex sialic acid type, and GNA-related lectins can bind with gp120 intensively, mainly due to their affinity for mannose and/or sialic acid [9, 10, 43, 44]. Accordingly, the combination of gp120 with GNA-related lectins instead of the receptor CD4 residing in host cells results in steric hindrance and prevents fusion and infection with HIV [4, 9, 43].

Similar to the anti-HIV mechanism of GNA-related lectins, we speculate that the binding between GNA-related lectins and sialic acid exposed on the glycoprotein receptors of host cells would prevent influenza A virus infection, mainly by blocking the combination of virus envelope glycoprotein HA with its corresponding sialic acid-linked receptor. Thus, GNA-related lectins that possess affinity for sialic acid would exert anti-influenza effects.

The antiviral activities of GNA-related lectins are realized by preventing the recognition between sialic acid and HA [45–47], whereas zanamivir and oseltamivir (two anti-H5N1 drugs) act mainly through the inhibition of the sialidase (also known as NA) activity of influenza A virus. NA plays a crucial role in cleaving sialic acid residues, and assists the release of progeny virions from infected cells [46]. Thus, the inhibition of sialidase activity could prevent progeny virion release, suppressing influenza A virus infection.

Some reports have revealed that HAs must accumulate mutations rapidly and continuously to escape recognition by the immune system [48, 49]. As the influenza A virus has mutated, it has affected a range of species from avians to humans. Avians are mainly infected by α 2,3-linked sialic acid specific influenza A virus strains, while humans are mainly infected by α 2,6-linked sialic acid specific influenza A virus strains [46, 50]. It has been suggested that the sialic acid affinity of HA may have remained unchanged during the evolution of this virus, despite the mutation of the HAs. If this is true, the anti-influenza activities of GNA-related lectins should be efficient and broad-spectrum.

In addition, GNA-related lectins may also exhibit the ability to selectively inhibit virus signaling pathways. As demonstrated in previous studies, the Ras-Raf and PI3K-Akt signaling pathways are the two of the most significant regulation mechanisms in virus propagation and vRNP (viral nucleoprotein) export [51–53]. The Ras-Raf signaling pathway appears to be particularly beneficial to virus replication, while blocking it results in strongly impaired growth of the influenza A virus [52, 53]. Additionally, the PI3K-Akt signaling pathway (considered to be a two-faced participant in virus defense and injection)—in contrast to most other antiviral signaling events—is not suppressed but rather activated in the presence of the viral NS1 protein at a very early stage in the replication cycle [52, 54, 55]. Nevertheless, recent research provides rational evidence that some GNA-related lectins, especially PCL, could induce murine fibrosarcoma L929 cell apoptosis and autophagy by negatively regulating the Ras-Raf and PI3K-Akt signaling pathways [56]. Here, GNA-related lectins could obviously act as antagonists in intracellular signaling pathways, inducing apoptosis and/or autophagy of cells infected with influenza A by blocking the Ras-Raf and PI3K-Akt pathways. However, more genomics and proteomics studies investigating this potential mech-

anism of GNA-related lectins are needed to support this viewpoint.

Conclusions

In summary, our results demonstrate that YFL-I appears to have the most potential as an anti-influenza agent due to its relatively strong affinity for sialic acid and its stable conformation when in complex with sialic acid. Other GNA-related lectins with similar properties to YFL-I may also exert anti-influenza activities by competitively blocking the combination of H1N1 virus envelope glycoprotein HA with its corresponding sialic acid-linked receptor in the host cell. Elucidating the anti-influenza effects of GNA-related lectins would lead to deeper investigations of the influenza A virus and other viruses that show sialic acid specificity, which would in turn allow potential candidate agents from the GNA-related lectin family against various sialic acid specific virus infections to be identified.

Acknowledgments We are grateful to Miss Mingwei Min (University of Cambridge) and Qian Liu (National University of Singapore) for their critical reviews of this manuscript. This work was supported in part by grants from the National Natural Science Foundation of China (General Programs: no. 30670469 and no. 30970643).

References

1. Goldstein IJ, Hughes RC, Monsigny T, Osawa T, Sharon N (1980) What should be called a lectin? *Nature* 285:66
2. Sharon N, Lis H (1989) Lectins as cell recognition molecules. *Science* 246:227–234
3. Liu B, Bian HJ, Bao JK (2010) Plant lectins: potential antineoplastic drugs from bench to clinic. *Cancer Lett* 287:1–12
4. Balzarini J (2007) Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy. *Nat Rev Microbiol* 5:583–597
5. Liu B, Zhang B, Min MW, Bian HJ, Chen LF, Liu Q, Bao JK (2009) Induction of apoptosis by *Polygonatum odoratum* lectin and its molecular mechanisms in murine fibrosarcoma L929 cells. *Biochim Biophys Acta* 1790:840–844
6. Liu B, Cheng Y, Zhang B, Bian HJ, Bao JK (2009) *Polygonatum cyrtoneuma* lectin induces apoptosis and autophagy in human melanoma A375 cells through a mitochondria-mediated ROS-p38-p53 pathway. *Cancer Lett* 275:54–60
7. Liu B, Cheng Y, Bian HJ, Bao JK (2009) Molecular mechanisms of *Polygonatum cyrtoneuma* lectin-induced apoptosis and autophagy in cancer cells. *Autophagy* 5:253–255
8. Li CY, Meng L, Liu B, Bao JK (2009) *Galanthus nivalis* agglutinin (GNA)-related lectins: traditional proteins, burgeoning drugs? *Curr Chem Biol* 3:324–333
9. An J, Liu JZ, Wu CF, Li J, Dai L, van Damme E, Balzarini J, De Clercq E, Chen F, Bao JK (2006) Anti-HIV I/II activity and molecular cloning of a novel mannose/sialic acid-binding lectin from rhizome of *Polygonatum cyrtoneuma* Hua. *Acta Biochim Biophys Sin (Shanghai)* 38:70–78
10. Tian Q, Wang W, Miao C, Peng H, Liu B, Leng FW, Dai L, Chen F, Bao JK (2008) Purification, characterization and molecular

- cloning of a novel mannose-binding lectin from rhizomes of *Ophiopogon japonicus* with antiviral and antifungal activities. *Plant Sci* 175:877–884
11. Liu J, Stevens DJ, Haire LF, Walker PA, Coombs PJ, Russell RJ, Gamblin SJ, Skehel JJ (2009) Structures of receptor complexes formed by hemagglutinins from the Asian Influenza Pandemic of 1957. *Proc Natl Acad Sci USA* 106:17175–17180
 12. Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, Vasisht N, Steinhauer DA, Daniels RS, Elliot A, Wiley DC, Skehel JJ (2004) The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* 303:1838–1842
 13. Lin T, Wang G, Li A, Zhang Q, Wu C, Zhang R, Cai Q, Song W, Yuen KY (2009) The hemagglutinin structure of an avian H1N1 influenza A virus. *Virology* 392:73–81
 14. Russell RJ, Stevens DJ, Haire LF, Gamblin SJ, Skehel JJ (2006) Avian and human receptor binding by hemagglutinins of influenza A viruses. *Glycoconj J* 23:85–92
 15. Skehel JJ, Wiley DC (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* 69:531–569
 16. Cohen J (2010) Swine flu pandemic. What's old is new: 1918 virus matches 2009 H1N1 strain. *Science* 327:1563–1564
 17. Chandrasekaran A, Srinivasan A, Raman R, Viswanathan K, Raguram S, Tumpey TM, Sasisekharan V, Sasisekharan R (2008) Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat Biotechnol* 26:107–113
 18. Stevens J, Blixt O, Tumpey TM, Taubenberger JK, Paulson JC, Wilson IA (2006) Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* 312:404–410
 19. van Riel D, Munster VJ, de Wit E, Rimmelzwaan GF, Fouchier RA, Osterhaus AD, Kuiken T (2007) Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals. *Am J Pathol* 171:1215–1223
 20. Shinya K, Ebina M, Yamada S, Ono M, Kasai N, Kawaoka Y (2006) Avian flu: influenza virus receptors in the human airway. *Nature* 440:435–436
 21. Nicholls JM, Bourne AJ, Chen H, Guan Y, Peiris JS (2007) Sialic acid receptor detection in the human respiratory tract: evidence for widespread distribution of potential binding sites for human and avian influenza viruses. *Respir Res* 8:73
 22. van Damme EJ, Nakamura-Tsuruta S, Smith DF, Ongenaert M, Winter HC, Rougé P, Goldstein IJ, Mo H, Kominami J, Cullerier R, Barre A, Hirabayashi J, Peumans WJ (2007) Phylogenetic and specificity studies of two-domain GNA-related lectins: generation of multispecificity through domain duplication and divergent evolution. *Biochem J* 404:51–61
 23. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612
 24. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* 34:W362–W365
 25. Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29:291–325
 26. Wright CS, Hester G (1996) The 2.0 Å structure of a cross-linked complex between snowdrop lectin and a branched mannopentaose: evidence for two unique binding modes. *Structure* 4:1339–1352
 27. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID (2009) DOCK 6: combining techniques to model RNA–small molecule complexes. *RNA* 15:1219–1230
 28. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701–1718
 29. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, RP SW, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide. Hochschulverlag AG, Zurich, ISBN 3 7281 2422 2
 30. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. *Intermolecular Forces* 11:331–342
 31. Schüttelkopf AW, van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr D* 60:1355–1363
 32. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an $N \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092
 33. Essmann U, Perera L, Berkowitz ML (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8592
 34. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472
 35. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3584–3590
 36. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38
 37. Liu B, Peng H, Yao Q, Li J, Van Damme E, Balzarini J, Bao JK (2009) Bioinformatics analyses of the mannose-binding lectins from *Polygonatum cyrtonema*, *Ophiopogon japonicus* and *Liparis noversa* with antiproliferative and apoptosis-inducing activities. *Phytomedicine* 16:601–608
 38. Liu B, Xu XC, Cheng Y, Huang J, Liu YH, Liu Z, Min MW, Bian HJ, Chen J, Bao JK (2008) Apoptosis-inducing effect and structural basis of *Polygonatum cyrtonema* lectin and chemical modification properties on its mannose-binding sites. *BMB Rep* 41:369–375
 39. Ding J, Bao J, Zhu D, Zhang Y, Wang DC (2010) Crystal structures of a novel anti-HIV mannose-binding lectin from *Polygonatum cyrtonema* Hua with unique ligand-binding property and super-structure. *J Struct Biol* 171:309–317
 40. Okimoto N, Futatsugi N, Fuji H, Suenaga A, Morimoto G, Yanai R, Ohno Y, Narumi T, Taiji M (2009) High-performance drug discovery: computational screening by combining docking and molecular dynamics simulations. *PLoS Comput Biol* 5:e1000528
 41. Newhouse EI, Xu D, Markwick PR, Amaro RE, Pao HC, Wu KJ, Alam M, McCammon JA, Li WW (2009) Mechanism of glycan receptor recognition and specificity switch for avian, swine, and human adapted influenza virus hemagglutinins: a molecular dynamics perspective. *J Am Chem Soc* 131:17430–17442
 42. Alonso H, Bliznyuk AA, Gready JE (2006) Combining docking and molecular dynamic simulations in drug design. *Med Res Rev* 26:531–568
 43. Balzarini J (2006) Inhibition of HIV entry by carbohydrate-binding proteins. *Antivir Res* 71:237–247
 44. Balzarini J, Schols D, Neyts J, van Damme E, Peumans W, de Clercq E (1991) Alpha-(1–3)- and alpha-(1–6)-D-mannose-specific plant lectins are markedly inhibitory to human immunodeficiency virus and cytomegalovirus infections in vitro. *Antimicrob Agents Chemother* 35:410–416
 45. Chand P, Bantia S, Kotian PL, El-Kattan Y, Lin TH, Babu YS (2005) Comparison of the anti-influenza virus activity of cyclopentane derivatives with oseltamivir and zanamivir in vivo. *Bioorg Med Chem* 13:4071–4077
 46. von Itzstein M (2007) The war against influenza: discovery and development of sialidase inhibitors. *Nat Rev Drug Discov* 6:967–974
 47. Yu K, Luo C, Qin G, Xu Z, Li N, Liu H, Shen X, Ma J, Wang Q, Yang C, Zhu W, Jiang H (2009) Why are oseltamivir and zanamivir effective against the newly emerged influenza A virus (A/H1N1)? *Cell Res* 19:1221–1224

48. Wang CC, Chen JR, Tseng YC, Hsu CH, Hung YF, Chen SW, Chen CM, Khoo KH, Cheng TJ, Cheng YS, Jan JT, Wu CY, Ma C, Wong CH (2009) Glycans on influenza hemagglutinin affect receptor binding and immune response. *Proc Natl Acad Sci USA* 106:18137–18142
49. Xu R, Ekiert DC, Krause JC, Hai R, Crowe JE Jr, Wilson IA (2010) Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. *Science* 328:357–360
50. Shriver Z, Raman R, Viswanathan K, Sasisekharan R (2009) Context-specific target definition in influenza a virus hemagglutinin–glycan receptor interactions. *Chem Biol* 16:803–814
51. Ludwig S, Planz O, Pleschka S, Wolff T (2003) Influenza-virus-induced signaling cascades: targets for antiviral therapy? *Trends Mol Med* 9:46–52
52. Ludwig S, Planz O, Pleschka S, Wolff T (2008) Signaling to life and death: influenza viruses and intracellular signal transduction cascades. In: Klenk HD, Matrosovich MN, Stech J (eds) *Avian influenza*. Basel, Karger, pp 210–224
53. Ludwig S (2009) Targeting cell signalling pathways to fight the flu: towards a paradigm change in anti-influenza therapy. *J Antimicrob Chemother* 64:1–4
54. Ehrhardt C, Wolff T, Pleschka S, Planz O, Beermann W, Bode JG, Schmolke M, Ludwig S (2007) Influenza A virus NS1 protein activates the PI3K/Akt pathway to mediate antiapoptotic signaling responses. *J Virol* 81:3058–3067
55. Ehrhardt C, Ludwig S (2009) A new player in a deadly game: influenza viruses and the PI3K/Akt signalling pathway. *Cell Microbiol* 11:863–871
56. Liu B, Wu JM, Li J, Liu JJ, Li WW, Li CY, Xu HL, Bao JK (2010) *Polygonatum cyrtoneuma* lectin induces murine fibrosarcoma L929 cell apoptosis and autophagy via blocking Ras-Raf and PI3K-Akt signaling pathways. *Biochimie* 92:1934–1938

Homology modeling, molecular dynamics, e-pharmacophore mapping and docking study of Chikungunya virus nsP2 protease

Kh. Dhanachandra Singh · Palani Kirubakaran · Shanthi Nagarajan · Sugunadevi Sakkiah · Karthikeyan Muthusamy · Devadasan Velmurgan · Jeyaraman Jeyakanthan

Received: 3 November 2010 / Accepted: 9 February 2011 / Published online: 29 March 2011
© Springer-Verlag 2011

Abstract To date, no suitable vaccine or specific antiviral drug is available to treat Chikungunya viral (CHIKV) fever. Hence, it is essential to identify drug candidates that could potentially impede CHIKV infection. Here, we present the development of a homology model of nsP2 protein based on the crystal structure of the nsP2 protein of Venezuelan equine encephalitis virus (VEEV). The protein modeled was optimized using molecular dynamics simulation; the junction peptides of a nonstructural protein complex were then docked in order to investigate the possible protein–protein interactions between nsP2 and the proteins cleaved by nsP2. The modeling studies conducted shed light on the

binding modes, and the critical interactions with the peptides provide insight into the chemical features needed to inhibit the CHIK virus infection. Energy-optimized pharmacophore mapping was performed using the junction peptides. Based on the results, we propose the pharmacophore features that must be present in an inhibitor of nsP2 protease. The resulting pharmacophore model contained an aromatic ring, a hydrophobic and three hydrogen-bond donor sites. Using these pharmacophore features, we screened a large public library of compounds (Asinex, Maybridge, TOSLab, Binding Database) to find a potential ligand that could inhibit the nsP2 protein. The compounds that yielded a fitness score of more than 1.0 were further subjected to Glide HTVS and Glide XP. Here, we report the best four compounds based on their docking scores; these compounds have IDs of 27943, 21362, ASN 01107557 and ASN 01541696. We propose that these compounds could bind to the active site of nsP2 protease and inhibit this enzyme. Furthermore, the backbone structural scaffolds of these four lead compounds could serve as building blocks when designing drug-like molecules for the treatment of Chikungunya viral fever.

Kh. Dhanachandra Singh, Palani Kirubakaran, and Shanthi Nagarajan contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1018-3) contains supplementary material, which is available to authorized users.

K. D. Singh · P. Kirubakaran · K. Muthusamy · J. Jeyakanthan (✉)
Department of Bioinformatics, Alagappa University,
Karaikudi 63003, India
e-mail: jkkanthan@gmail.com

S. Nagarajan
Bioinformatics Centre, Pondicherry University,
Pondicherry 605 014, India

S. Sakkiah
Bioinformatics Centre and Computational Biology Laboratory,
Department of Bioinformatics, Bharathiar University,
Coimbatore 641 046, India

D. Velmurgan
CAS, Biophysics and Crystallography, University of Madras,
Chennai 600025, India
e-mail: d_velu@yahoo.com

Keywords Chikungunya virus · Homology modeling · Desmond · Nonstructural proteins (nsP) · Molecular dynamics simulation (MDS)

Introduction

Chikungunya virus (CHIKV) is a member of the genus *Alphavirus* of the family *Togaviridae*. While CHIKV is transmitted to humans by several species of mosquitoes, *A. aegypti* and *A. albopictus* are the two main vectors [1]. In

Africa and Asia, two outbreaks of Chikungunya virus have occurred, with time intervals of 7–8 years to 20 years between consecutive epidemics [2]. In recent years, emerging and re-emerging tropical infectious diseases have been shown to have high social and economic impacts [3]. The symptoms of CHIKV infection are fever, headache, nausea, vomiting, myalgia, rash, and arthralgia [4]. When infected, patients are observed to have painful puffy feet and ankles, and they experience predominantly chronic polyarthralgia, typically a rheumatoid arthritis-like illness [5]. Recurrent episodes can occur in some patients, generally in the form of chronic and persistent arthralgia. The CHIKV genome consists of a linear, positive-sense, single-stranded RNA molecule of about 11.8 kb [6]. Alphaviruses produce two mRNAs after infection, such as genomic (49 S) RNA, which is translated into nonstructural (replicase) proteins, and subgenomic (26 S) RNA, which serves as mRNA for virion structural proteins. Nonstructural proteins (nsP) such as nsP1, nsP2, nsP3 and nsP4 are formed by proteolytic cleavage of a long polyprotein that includes individual nonstructural proteins [7].

Alphaviruses are enveloped particles, and their genome consists of a single-stranded positive-sense RNA molecule of approximately 12,000 nucleotides. The genome of CHIKV is considered to have the following gene arrangement: 5' cap-nsP1-nsP2-nsP3-nsP4-(junction region)-C-E3-E2-6 K-E1-poly(A) 3' [8]. Coding sequences of CHIKV from Indian Ocean patients consisted of two open reading frames (ORFs) encoding the nonstructural proteins (consisting of nsP1, 2, 3 and 4) in the form of two polyprotein precursors in the 5' two-thirds of the genome, and the other one-third containing one ORF encoding the structural polyprotein (consisting of C, E3, E2 and E1) [9].

The alphavirus nsP2 protease domain belongs to the papain superfamily of cysteine proteases. The proteolytic function of Semliki Forest virus (an alphavirus) nsP2 has been mapped to its C-terminal domain [10]. The nsP2 contains the protease domain responsible for nsP maturation [11]; hence, it is clear that the nsP2 is essential for alphavirus replication and exhibits some degree of sequence specificity among alphaviruses [12], making it an attractive target for broad-spectrum antiviral inhibitor development. The nsP2 (799 amino acids) protein from Semliki Forest virus and Sindbis virus (SINV) specifically cleaves the γ , β -triphosphate bond at the 5' end of RNA. This activity is restricted to the N-terminal domain, and the C-terminal domain has no RNA triphosphatase activity [13].

It is necessary to understand the structural aspects of CHIKV nsP2 and other alphavirus nsP2 proteins in order to raise broad-spectrum viral inhibitors. In this report, we explain the structural features of the nsP2 protein, which is modeled based on the crystal structure of Venezuelan equine encephalitis virus (VEEV, PDB ID: 2HWK).

Molecular dynamics simulations (MDs) provide an alternative tool for biological problems that are complementary to experimental techniques [14]. The modeled nsP2 protein structures were refined using MD techniques, and exhaustive docking analysis was performed using induced fit docking [15, 16], which is a robust automated docking method that predicts the conformations of flexible ligands bound to macromolecular targets. Three peptides, (A) Gly1-Ala2-Gly3-Ile4, (B) Gly1-Cys2-Ala3-Pro4 and (C) Gly1-Gly2-Trp3-Ile4, were docked. The plausible docking poses obtained provide a great deal of information about the protein-cleaving mechanism of nsP2 protease. The critical residues in nsP2 were identified by docking three different peptides in order to identify the residues responsible for nonstructural protein cleavage. We propose five different pharmacophore sites based on the receptor-peptide interactions. The structural and pharmacophore information can be effectively used for broad-spectrum inhibitor identification, which, Chikungunya virus aside, should prove useful considering that alphavirus family candidates are being used as bioweapons.

Materials and methods

All computational analysis was carried out on a Red Hat 5.3 Linux platform running on a Lenovo PC with an Intel Core 2 Duo processor and 2 GB of RAM.

Homology model development

Homology modeling is a theoretical method that is used to predict the structure of a sequence with an accuracy that is comparable to the best results achieved experimentally. The modeled protein quality is extremely dependent on the identity between the target and template proteins. The nonstructural polyprotein (nsP) of Chikungunya virus (strain S27, African prototype) has 2,470 amino acids and contains four chains, namely nsP1 (1–535), nsP2 (536–1333), nsP3 (1334–1863), and nsP4 (1864–2474). The nsP2 protease is an essential protein whose proteolytic activity is critical for virus replication. On the other hand, experimentally derived protein information is unavailable, so we decided to model the Chikungunya virus nsP2 protease using the structure of its closest homolog that has been solved by crystallography.

The nsP2 protein sequence was collected from the Swiss-Prot Protein Database (accession number: Q8JUX6) [8, 17]. A similarity search for nsP2 protease in the Protein Data Bank (<http://www.rcsb.org>) was performed using the BLAST server [18]. The protein similarity search identified a very similar protein structure belonging to the nsP2 protease of the alphavirus Venezuelan equine encephalitis

(VEE) [19] (PDB ID: 2HWK), which has 40% sequence identity with CHIKV nsP2, so this structure was used as a template to generate the model. This was the only template available to model the nsP2 protein. The model was generated using Prime (Schrodinger, LLC, New York, USA) [20], and then the energy was minimized using the OPLS (optimized potentials for liquid simulations) 2005 force-field [21].

Model validation

The validation of the structure model obtained from Prime was performed by inspecting the psi/phi Ramachandran plot obtained from PROCHECK analysis [22]. The PROSA [23] test was applied to the final model to check energy criteria against the potential of mean force derived from a large set of known protein structures. The root mean square deviation (RMSD) between the main chain atom of the model and the template was calculated by superimposing the structure of the template (2HWK) on the predicted structure of CHIKV nsP2 protease in order to assess the reliability of the model using PyMol [24].

Molecular dynamics simulations

All molecular dynamics (MD) simulations were performed using the program Desmond [15], which uses a particular “neutral territory” method called the midpoint method [25] to efficiently exploit a high degree of computational parallelism. The OPLS 2005 force-field [21] was used to model the amino acid interactions in the protein, and the SPC (simple point charge) method [26] was used for the water model. Equilibration of the system was carried out using the default protocol provided in Desmond, which consists of a series of restrained minimizations and molecular dynamics simulations that are designed to slowly relax the system without deviating substantially from the initial protein coordinates. The initial coordinates for the MD calculations were taken from the modeled protein. The SPC water molecules were then added (the orthorhombic dimensions of each water box were $10 \text{ \AA} \times 10 \text{ \AA} \times 10 \text{ \AA}$ approximately, which ensured that the whole surfaces of the complexes were covered), and the system was neutralized by adding Cl^- counterions to balance the net charge of the system. After the construction of the solvent environment, each complex system was composed of about 34,942 atoms. Before equilibration and the long production MD simulations, the systems were minimized and pre-equilibrated using the default relaxation routine implemented in Desmond. The whole system was subjected to 300 K for 5 ns of simulation, and the final conformations of the modeled protein are presented below. The structural changes and dynamic behavior of the protein were analyzed by calculating the RMSD and energy.

Active site predictions

The active site of the modeled protein was investigated using the SiteMap program [27]. This software generates information on the binding site's characteristics using novel search and analytical facilities: a SiteMap calculation begins with an initial search step that identifies or characterizes—through the use of grid points—one or more regions on the protein surface that may be suitable for binding ligands to the receptor. Contour maps are then generated, producing hydrophobic and hydrophilic maps [28]. The hydrophilic maps are further divided into donor, acceptor, and metal-binding regions. The evaluation stage, which concludes the calculation, involves assessing each site by calculating various properties: the number of site points, a measure of the size of the site; exposure/enclosure, two properties providing different measures of how available the site is to the solvent; contact, which measures how strongly the average site point interacts with the surrounding receptor via van der Waals nonbonding interactions; donor/acceptor character, a property related to the sizes and intensities of H-donor and H-acceptor regions; and SiteScore, an overall property based on the previous properties, constructed and calibrated so that the average SiteScore for a promising binding site is 1.0 [28].

Induced fit docking simulation of peptides

To keep the receptor flexible in the docking protocol, we used a mixed molecular docking/dynamics protocol called induced fit docking (IFD) [29], as developed by Schrödinger, LLC (<http://www.schrodinger.com/>). The IFD protocol used in this study was carried out in three consecutive steps [30]. First, the ligand was docked into a rigid receptor model with scaled-down van der Waals (vdW) radii. A vdW scaling of 0.5 was used for both the protein and ligand nonpolar atoms. A constrained energy minimization was carried out on the protein structure, keeping it close to the original crystal structure while removing bad steric contacts. Energy minimization was carried out using the OPLS 2005 force-field with an implicit solvation model until default criteria were met. The active site predicted in SiteMap was used to define the location of the binding site and the dimension of the energy grids for initial docking. The Glide XP mode was used for the initial docking, and 20 ligand poses were retained for protein structural refinement. In the second step, Prime was used to generate the induced-fit protein–ligand complexes. Each of the 20 structures from the previous step was subjected to side-chain and backbone refinements [30]. All residues with at least one atom located within 4.0 \AA of each corresponding ligand pose were included in the Prime refinement. The refined complexes were ranked by Prime energy, and the receptor structures

within 30 kcal mol⁻¹ of the minimum energy structure were put through to a final round of Glide docking and scoring. In the final step, each ligand was redocked into every refined low-energy receptor structure produced in the second step using Glide XP at default settings.

The binding modes of three different peptides that are believed to be cleaved by nsP2 were theoretically identified by docking calculations. The prime goal of the nsP2 protein is to cleave the nonstructural protein complex into individual active proteins. No crystal structures or appropriate templates are available to model the nsP1 and nsP4 proteins, so we made use of junction peptide models to explain the protein–protein interactions. The nsP2 protease cleaving sites are reported in the Swiss-Prot database; the cleavage site information was obtained by means of similarity analysis carried across alphavirus family members. In the database, two peptides are reported for every junction of nsP1–nsP2, nsP2–nsP3, and nsP3–nsP4. In the molecular docking studies, we used the junction peptides as substrates; moreover, we considered that the flanking residues provide access to the real system.

Energy-optimized pharmacophore mapping

Energy-optimized pharmacophores (e-pharmacophores) are obtained by mapping the energetic terms from the Glide XP scoring function onto atom centers. The ligand is docked with Glide XP and the pose is refined. The Glide XP scoring terms are computed, and the energies are mapped onto atoms. Next, pharmacophore sites are generated, and the Glide XP energies from the atoms that comprise each pharmacophore site are summed. The sites are then ranked based on these energies, and the most favorable sites are selected for the pharmacophore hypothesis. These pharmacophores are then used as queries for virtual screening [31].

Pharmacophore sites were automatically generated from the protein–ligand docked complex with Phase (Phase, v.3.0, Schrodinger, LLC) using the default set of six chemical features: hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobe (H), negative ionizable (N), positive ionizable (P), and aromatic ring (R). Phase treats most cationic groups as being exclusively positive ionizable. Hydrogen-bond acceptor sites were represented as vectors along the hydrogen bond axis in accordance with the hybridization of the acceptor atom. Hydrogen-bond donors were represented as projected points, located at the corresponding hydrogen-bond acceptor positions in the binding site. Projected points allow the possibility of structurally dissimilar active compounds forming hydrogen bonds to the same location, regardless of their point of origin and directionality [31].

Each pharmacophore feature site was first assigned an energetic value equal to the sum of the Glide XP

contributions of the atoms comprising the site. This allows sites to be quantified and ranked on the basis of these energetic terms. Glide XP descriptors include terms for hydrophobic enclosure, hydrophobically packed correlated hydrogen bonds, electrostatic rewards, π – π stacking, π ... cation, and other interactions [32]. Sites where less than half of the heavy atoms contribute to the pharmacophore feature were excluded from the final hypothesis. Thus, if only two heavy atoms in a six-membered ring exhibit energetic interactions, the ring is not considered a pharmacophore feature [31].

e-Pharmacophore database screening

For the e-pharmacophore approach, explicit matching was required for the most energetically favorable site, provided that it scored better than -1.0 kcal mol⁻¹. Multiple sites were included in cases where more than one site had the top score. Screening molecules were required to match a minimum of 3 sites for a hypothesis with 3 or 4 sites and a minimum of 4 sites for a hypothesis with 5 or more sites. The distance matching tolerance was set to 2.0 Å as a balance between stringent and loose-fitting alignment. Screening of compounds was performed against Asinex (<http://www.asinex.com>), Maybridge (<http://www.maybridge.com>), TOSLab (<http://www.toslab.com>), Binding Database (<http://www.bindingdb.org/bind/index.jsp>), etc. Database hits were ranked in order of fitness score, a measure of how well the aligned ligand conformer matches the hypothesis based on RMSD site matching, vector alignments, and volume terms. The fitness scoring function is an equally weighted composite of these three terms and ranges from 0 to 3, as implemented in the default database screening of Phase. The ligands were selected based on the fitness score. The ligands with the best fitness scores were docked into the binding sites of the modeled protein [31].

ADME screening

The QikProp program [33] was used to obtain the ADME properties of the compounds. This predicts both physically significant descriptors and pharmaceutically relevant properties. All of the compounds were neutralized before being used by QikProp. The neutralizing step is essential, as QikProp is unable to neutralize a structure and no properties will be generated in the normal mode. The program was processed in normal mode, and predicted 44 properties for the molecules, consisting of principal descriptors and physiochemical properties, along with a detailed analysis of log *P* (octanol/water), QP%, and log HERG. It also evaluated the acceptability of the compounds based on Lipinski's rule of five [34], which is essential for rational drug design.

Results and discussion

nsP2 model generation

The model was generated based on the nsP2 protease of Venezuelan equine encephalitis alphavirus (VEEV); this model has similar structural features to the template protein (Fig. 1). The RMSD between the minimized protein model and the template structure was found to be 1.1 Å (Fig. 2). The modeled protein was energy minimized using the OPLS 2005 force-field. Both domains are composed of α-helices and β-strands (Fig. 3). The N-terminus is dominated by an α-helix. The C-terminal domain contains helices and strands. The central β-sheets are flanked by α-helices. However, the function of the C-terminal domain of nsP2 is not clear [19]. The three-dimensional structure provides valuable insight into molecular function and also enables the protein–protein interaction to be analyzed.

Validation of the predicted structure

The overall stereochemical quality of the model was assessed by PROCHECK. The Ramachandran plot showed 81.6% of the residues in the most favorable region, 17.4% in the allowed region, 0.7% in the generously allowed region and 0.3% in the disallowed region; the corresponding values for the 2HWK template were 88.6%, 11.0%, 0.0% and 0.4%, respectively. These results revealed that the majority of the amino acids are in a phi-psi distribution that is consistent with a right-handed α-helix, and the model is reliable and of good quality (see Fig. 1 of the “Electronic supplementary material,” ESM). The G-factors, indicating the quality of the covalent, dihedral and overall bond angles, were −0.10° for dihedrals, 0.42° for covalent, and −0.11° overall. The overall main-chain and side-chain parameters, as evaluated by ProCheck, are all very favorable. The Ramachandran plot characteristic and G-factors confirm the quality of the predicted model. In order to investigate whether the

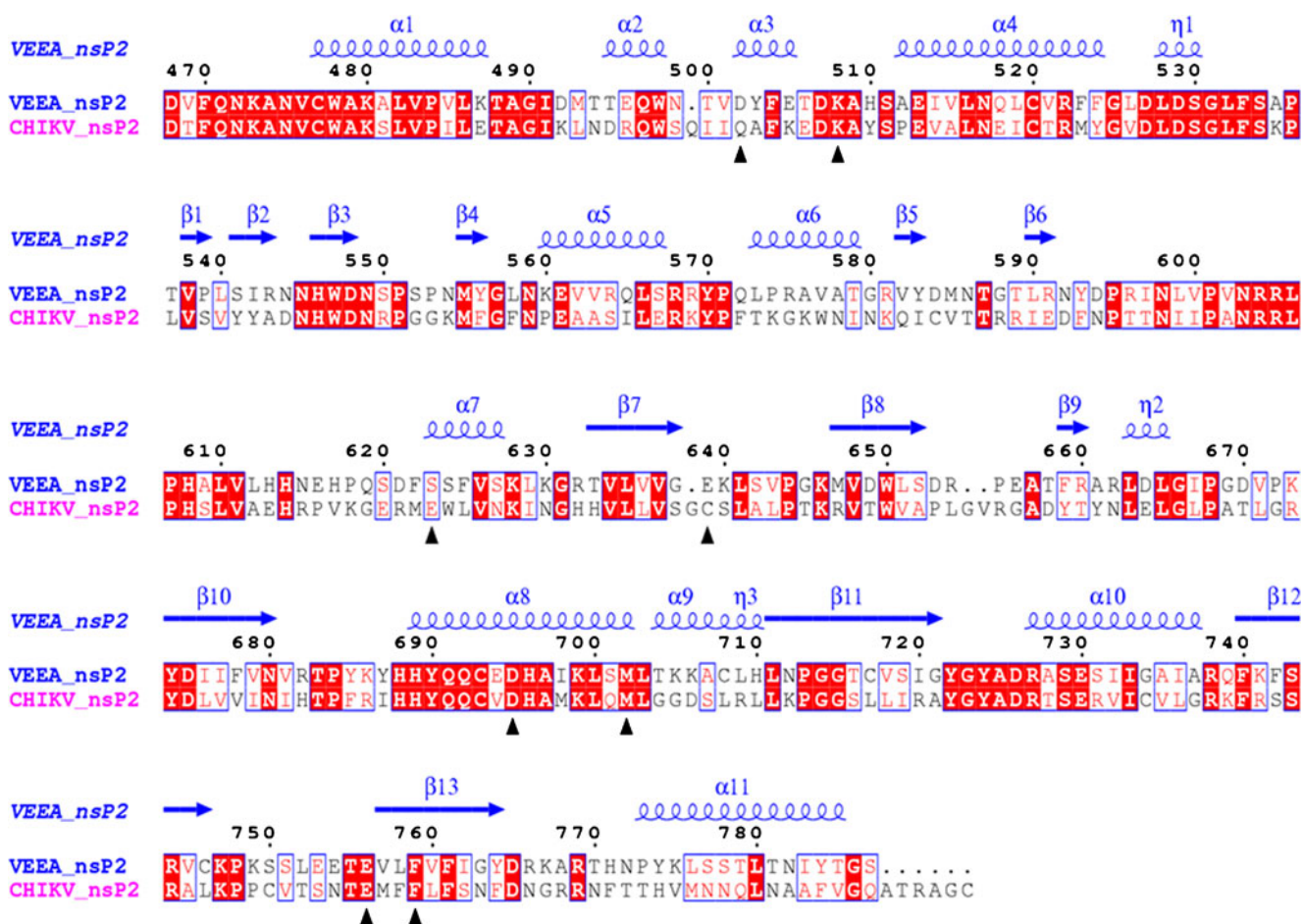


Fig. 1 Structure-based sequence alignment of VEEA_nsP2 (PDB ID: 2HWK) from Venezuelan equine encephalitis alphavirus nsP2 protease with the modeled Chikungunya virus nsP2 protease

(CHIKV_nsP2). The secondary structural elements are shown above the alignment. Active site and peptide interaction residues are highlighted with black triangles

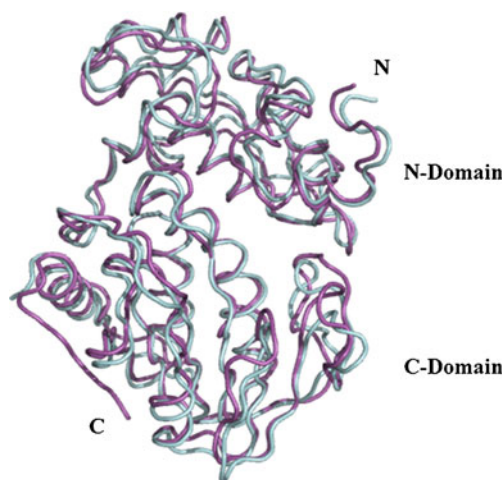


Fig. 2 Superimposition of template (*magenta*) and model protein (*cyan*)

interaction energy of each residue with the remainder of the protein is negative, a second test was done to apply energy criteria using a ProSA energy plot. The ProSA analysis of the model showed that almost all of the residues had negative interaction energies, with very few residues displaying positive interaction energies, as shown in Fig. 2 of the ESM.

Molecular dynamics simulations

Molecular dynamics can be used to explain protein structure–function problems, such as folding, conformational flexibility and structural stability. In the simulations, we monitored the backbone atoms and the C- α -helix of the modeled protein. The RMSD values of the modeled structure's backbone atoms were plotted as a time-dependent function of the MD simulation. The results support our modeled structure, as they show constant RMSD deviation throughout the whole simulation process. Graphs of potential energy, temperature, pressure and volume are shown in Fig. 3 of the ESM. The time dependence of the RMSD (\AA) of the backbone atoms of the modeled protein during a 5 ns simulation is shown in Fig. 4.

The graph clearly indicates that there is a change in the RMSD from 1.0 \AA to 3.0 \AA in the nsP2 homology model during the first 1500 ps, but after that it reaches a plateau. The RMSD values of the backbone atoms in the system tend to converge after 2000 ps, showing fluctuations of around 1 \AA . The low RMSD and the simulation time indicate that, as expected, the 3D structural model of nsP2 represents a stable folding conformation.

Induced fit docking (IFD) results for peptides

Molecular docking simulation helps us to understand the plausible binding modes and interactions. In IFD, we are

able to make the receptor flexible, which helps when refining the active sites. The active site of the protein lies in the C-terminal domain (Fig. 5). All three peptides interact with the C-terminal domain, except for the glycine of GAGI and the isoleucine of GGYI.

Gly534-Ala535-Gly536-Ile537 (GAGI) interaction

In total, GAGI forms six hydrogen bonds to the active site of the nsP2 protein (Fig. 6a). The Gly534 backbone nitrogen forms a trifurcated hydrogen bond with the oxygens of Ser1293, Glu1296 and Glu1157. However, in reality, the NH_2 at the starting position can form only two interactions, as it is attached to the preceding residue and hence should be a backbone NH. Ala535 does not interact with the protein residues. The Gly536 backbone oxygen forms a hydrogen bond with the nitrogen of Gln1039. The Ile537 side-chain oxygen and nitrogen form hydrogen bonds with the oxygen and nitrogen of His1222 and Lys1239.

Gly1332-Cys1333-Ala1334-Pro1335 (GCAP) interaction

GCAP also forms six hydrogen bonds (Fig. 6b), and all of the peptides form hydrogen bonds with the nsP2 protein. The NH_2 of Gly1332 forms a trifurcated H-bond with the oxygens of Ser1293, Glu1157 and Glu1297. However, in reality it can form only two bonds, as explained above. The backbone oxygen and nitrogen of Cys1333 and Ala1334 form hydrogen bonds with the nitrogen and oxygen side chains of His1222. The side-chain oxygen of Pro1335 forms a hydrogen bond with the nitrogen of Lys1045.

Gly1862-Gly1863-Tyr1864-Ile1865 (GGYI) interaction

GGYI forms seven hydrogen bonds with the nsP2 protein (Fig. 6c). Gly1862 forms a trifurcated H-bond with the

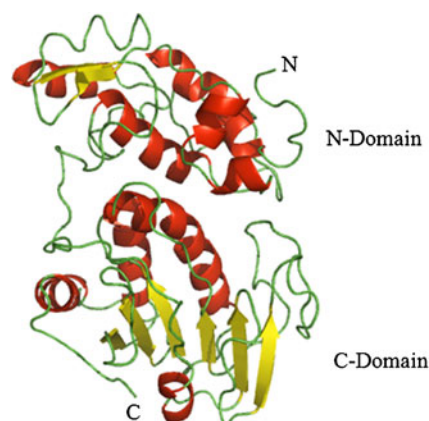
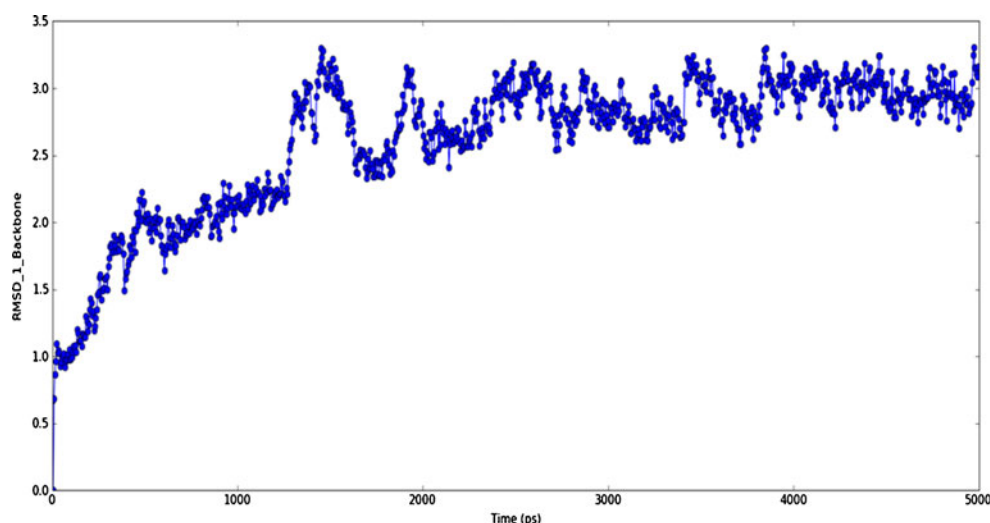


Fig. 3 Ribbon diagram of the modeled Chikungunya virus nsP2 protease (CHIKV_nsP2) showing the N- and C-terminal domains. α -Helices, β -strands and loops are colored *red*, *yellow* and *green*, respectively

Fig. 4 RMSD of the backbone atoms of the modeled protein over a time period of 5 ns



same residue, but only two H-bonds are possible in reality. The Tyr1864 backbone nitrogen and oxygen form hydrogen bonds with the oxygen of His1222 and nitrogen of Lys1239. The phenolic group of tyrosine forms a hydrogen bond with the oxygen of Gly1176. The backbone nitrogen of Ile1865 forms a hydrogen bond with the nitrogen of Lys1045. The Glide scores, Glide energies and IFD scores are shown in Table 1.

e-Pharmacophore development and screening of the database

The e-pharmacophore combines aspects of structure-based and ligand-based techniques. Incorporating protein–ligand contacts into ligand-based pharmacophore approaches has been shown to produce enhanced enrichments over using ligand information alone. The method described here attempts to take a step beyond simple contact scoring by

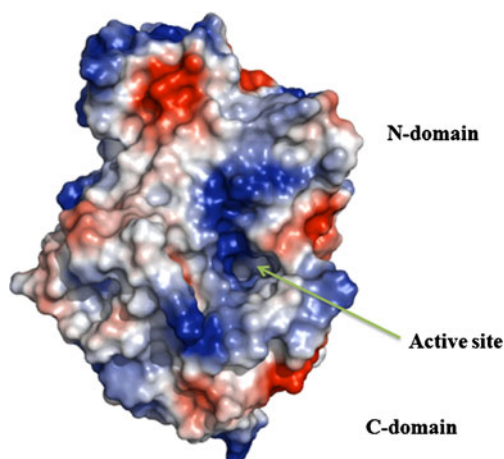


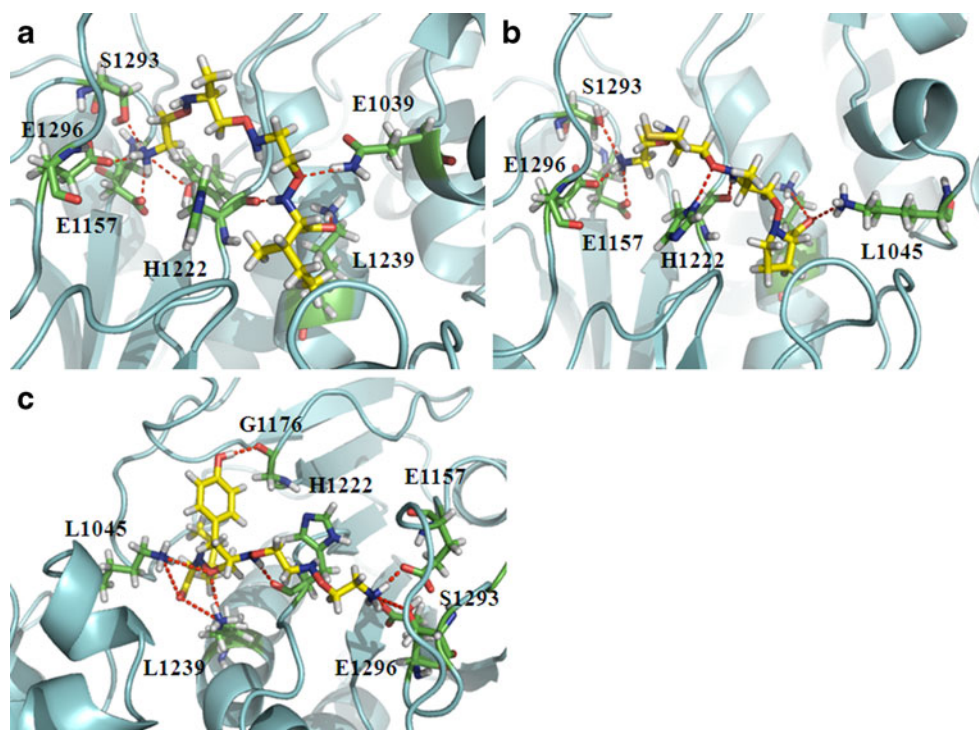
Fig. 5 Electrostatic potential surface of the CHIKV_{nsP2} and its active site pocket. The positively and negatively charged surface regions are shown in *blue* and *red*, respectively

incorporating structural and energetic information using the scoring function in Glide XP. Seven pharmacophore sites were predicted, but only five pharmacophore sites are chosen based on the score. The final hypothesis consists of a hydrophobic group (H), an aromatic ring (R), and three H-bond donors (D), and their distances are shown in Fig. 7a and b, respectively. The hydrophobic site H23 lies in the COOH group of Ile537 (GAGI), while the donor D10 lies in the phenol group of Tyr1864 (GGYI). The donor D12 lies in the amino group of Gly536 (GAGI), Cys1333 (GCAP) and Tyr1864 (GGYI), and the donor D16 lies in the amino group of Gly534 (GAGI) and Gly1862 (GGYI). These energetically favorable sites encompass the specific interactions of junction peptides and the nsP2 protein, and this information should prove helpful in the development of new nsP2 inhibitors. With this pharmacophore hypothesis, compound screening was performed against Asinex, Maybridge, Binding Database, TOS Lab, etc.: a total of more than 300,000 compounds. To further enrich the screening, excluded volumes were added to the hypotheses. Receptor-based excluded volumes were included in order to help reduce false positives by eliminating inactive compounds that cannot simultaneously match the hypothesis and avoid clashing with the receptor. Compounds with fitness scores of more than 1.0 were subjected to Glide high-throughput virtual screening (HTVS).

Glide extra precision docking

Molecular docking is a computational technique that samples conformations of small compounds at protein-binding sites; scoring functions are used to assess which of these conformations best complement the protein-binding site. There are two main aspects to assessing the quality of docking methods: (i) docking accuracy, which recognizes the true binding mode of the ligand to the target protein,

Fig. 6 Binding modes of the junction peptides Gly534-Ala535-Gly536-Ile537 (a), Gly1332-Cys1333-Ala1334-Pro1335 (b) and Gly1862-Gly1863-Tyr1864-Ile1865 (c), based on induced fit docking results



and (ii) screening enrichment, which measures how much better a docking method is at identifying true binding ligands than random screening.

To speed up the screening of a large set of compound databases, we took the refined active site from IFD and made

this active site rigid while screening in HTVS and Glide XP docking. Twenty compounds were selected from the HTVS for further Glide XP docking study based on the Glide score. Here, we report the four compounds from this Glide XP docking study with the best Glide scores (−7.5 to 9.8) and

Table 1 The peptides used in the docking simulations and their corresponding Glide scores and Glide energies

Peptide	Junction	Glide score	Glide energy (kcal/mol)	IFD score ^a	Interaction (D...H–A) ^b	H-bond length (Å)
Gly534-Ala535-Gly536-Ile537 (GAGI)	nsP1–nsP2	−7.538	−48.743	−631.584	NH...O Ser1293)	1.830
					NH...O(Glu1296)	1.833
					NH...O(Glu1157)	1.649
					NH(Gln1039)...O	2.008
					NH(Lys1239)...O	2.073
					NH...O(His1222)	1.828
Gly1332-Cys1333-Ala1334-Pro1335 (GCAP)	nsP2–nsP3	−7.399	−46.500	−631.359	NH...O(Ser1293)	1.983
					NH...O(Glu1157)	1.643
					NH...O(Glu1296)	2.037
					NH(His1222)...O	1.914
					NH...O(His1222)	1.712
Gly1862-Gly1863-Tyr1864-Ile1865 (GCYI)	nsP3–nsP4	−12.045	−59.956	−635.896	NH(Lys1045)...O	2.107
					NH...O(Ser1293)	2.454
					NH...O(Glu1157)	1.699
					NH...O(Glu1296)	2.138
					NH...O(His1222)	1.820
					OH(Gly1176)...O	1.931
NH(Lys1239)...O	2.228					
NH(Lys1045)...N	2.039					

^a IFD score induced fit docking score

^b D donor, H hydrogen, A acceptor

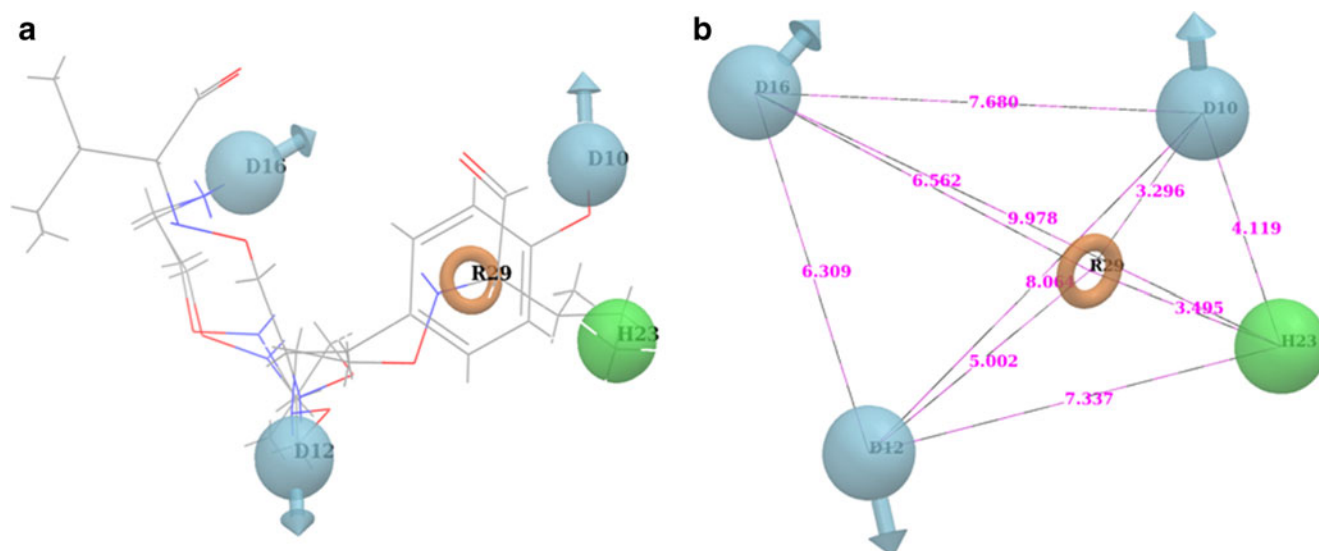


Fig. 7 Common pharmacophore hypothesis (DDDDR) based on the alignment of junction peptides (a) and the distance between the pharmacophore sites (b). Geometry of the pharmacophore. Orange torus aromatic ring feature, green sphere hydrophobic feature, light blue sphere donor feature

Glide energies (-29 to -49 kcal mol $^{-1}$), which suggest strong enzyme–ligand interactions. The chemical names of the four lead compounds and their corresponding database identity (ID) numbers are: [5-(5-fluoro-2,4-dioxypyrimidin-1-yl)-3,4-dihydroxoxolan-2-yl]methyl phosphate (27943: Binding Database); 4-[hydroxy-(4-methylphenyl)methylidene]-5-(3-hydroxyphenyl)-1-(2-hydroxypropyl)pyrrolidine-2,3-dione (21362: TosLab); *N*-(2-methyl-4-nitrophenyl)-2-[[5-(6-oxocyclohexa-2,4-dien-1-ylidene)-1,2-dihydro-1,2,4-triazol-3-yl]sulfanyl]acetamide (ASN 01107557: Asinex); and *N*-(5-

ethyl-1,3,4-thiadiazol-2-yl)-2-[[5-(6-oxocyclohexa-2,4-dien-1-ylidene)-1,2-dihydro-1,2,4-triazol-3-yl]sulfanyl]acetamide (ASN 01541696: Asinex). The chemical structures of these lead molecules are illustrated in Fig. 8, and the binding modes of these four lead molecules and their interacting residues are shown in Fig. 9a–d and Table 2. The residues Lys1045, Gly1176, His1222 and Lys1239 are involved in ligand interactions and are also important residues for the peptide interaction. The nsP2 protease sequences available in UniProt were retrieved (UniProt IDs: Q8JUX6, D7R997,

Fig. 8 Structures of the four lead molecules along with their compound IDs and database names

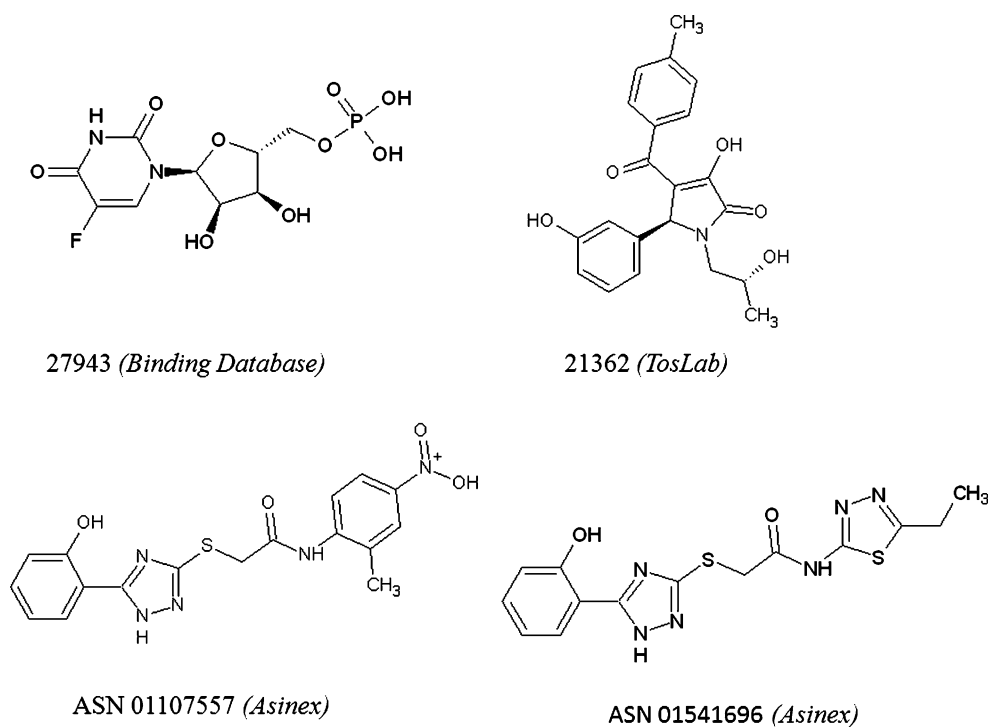
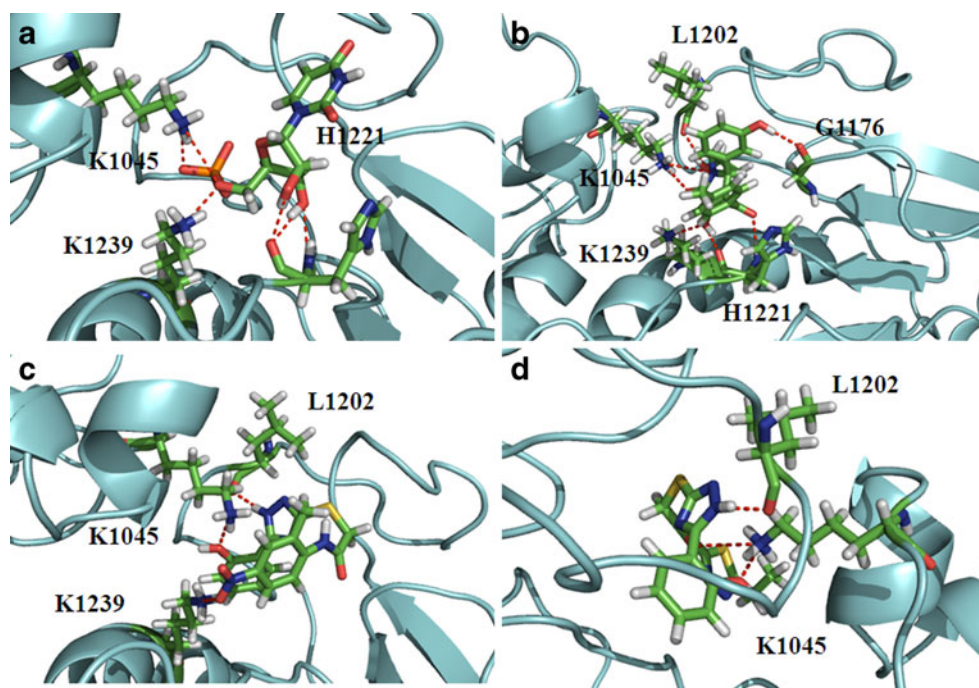


Fig. 9 Binding modes of the four potential ligands to the active site of nsP2 protease. The compound database IDs of the lead molecules are as follows: 27943: Binding Database (a), 21362: TOSLab (b), ASN 01107557: Asinex (c), ASN 01541696: Asinex (d)



D7R977, D7R966, D7R942, D7R973, A6MH22, D7R926, D7R999, D7R979, A0FJ31, D7R944, D7R9A1, Q1H8W7, Q1EL94, C7G0W2, A6MH12, B4YIR3, A9LMA5, A5Y7Y4, D2KBP9, C7AE71, Q1W368, D2KBQ5, D2CY36 and C7ADY7) and multiple sequence alignment was performed in ClustalW. We found that there were no changes in the residues at the active site we reported. Multiple sequence alignment of the reported sequences for nsP2 protease is shown in Fig. 4 of the ESM. Hence, the reported active site does not change very often.

Predicted ADME properties

We analyzed 44 physically significant descriptors [33] and pharmaceutically relevant properties of the four lead compounds, including molecular weight, H-bond donors, H-bond acceptors, $\log P$ (octanol/water), $\log P$ MDCK, $\log K_p$ (skin permeability), humoral absorption, and their positions according to Lipinski's rule of five (Table 3 and Table 4). Lipinski's rule of five is a rule of thumb to evaluate drug likeness; in other words, to determine if a

Table 2 Pharmacophore results for the best hit compounds, and Glide XP docking results

Lead molecules ^a	Interaction (D...H-A) ^b	H-bond length (Å)	Align score	Fitness score	Glide score	Glide energy (kcal/mol)
27943	OH... O(His1222)	1.982	1.015	1.029	-9.601	-42.576
	OH... O(His1222)	1.932				
	NH(Lys1239)...O	1.766				
21362	OH...O(Gly1176)	1.974	0.927	1.046	-9.279	-48.690
	NH(His1222)...O	2.161				
	OH...O(His1222)	1.621				
	NH(Lys1239)...O	1.706				
	NH(Lys1045)...O	1.893				
	OH...O(Leu1203)	2.156				
ASN 01107557	NH(Lys1045)...O	1.833	0.981	1.170	-8.57	-32.436
	NH...O(Leu1203)	1.812				
	NH...O(Lys1239)	2.956				
ASN 01541696	NH...O(Leu1202)	1.848	1.240	1.009	-7.640	-29.999
	NH(Lys1045)...O	1.790				
	NH(Lys1045)...O	2.707				

^a Ligand IDs are: 43077: Binding Database; 21362: TOSLab; ASN 01107557 and ASN 01541696: Asinex database

^b D donor, H hydrogen, A acceptor

Table 3 Principal descriptors calculated by Qikprop simulation

Lead molecules ^a	Molecular weight ^b (g/mol)	Molecular volume ^c (Å ³)	PSA ^d	HB donors ^e	HB acceptors ^f	Rotatable bonds ^g
27943	342.174	851.553	192.165	5.000	13.600	3
21362	367.401	1145.733	118.162	2.000	7.200	4
ASN 01107557	385.397	1154.389	139.73	3.000	6.750	4
ASN 01541696	362.424	1099.505	130.068	3.000	7.750	5

^a Ligand IDs: 43077: Binding Database; 21362: TOSLab; ASN 01107557 and ASN 01541696: Asinex database

^b Molecular weight of the molecule

^c Total solvent-accessible volume in cubic angstroms using a probe with a radius of 1.4 Å

^d Van der Waals surface areas of polar nitrogen and oxygen atoms

^e Estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer

^f Estimated number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer

^g Number of rotatable bonds

chemical compound with a certain pharmacological or biological activity has properties that would likely make it an orally active drug in humans. The rule describes molecular properties that are important in the drug's pharmacokinetics in the human body, including its ADME. However, the rule does not predict whether a compound is pharmacologically active. The four selected compounds were in the acceptable range of Lipinski's rule of five. For the four lead compounds, the partition coefficient ($QP \log P(o/w)$) and the water solubility ($QP \log S$), which are crucial when estimating the absorption and distribution of drugs within the body, ranged between -1.690 to 1.727 and -1.582 to -4.691 , respectively, while the cell permeability ($QP PCaco$), a key factor governing drug metabolism and its access to biological membranes, ranged from 0.345 to 95 . Overall, the percentage human oral absorptions for the compounds ranged from 25% to 100% .

All of these pharmacokinetic parameters are within the acceptable range defined for human use, thereby indicating their potential for use as drug-like molecules.

Conclusions

Our main objective of this work was to identify the residues involved in the cleavage mechanism through theoretical calculations. The identification of inhibitors for Chikungunya virus has been hampered but a lack of structural insight into any proteins. Therefore, we have chosen to model the nsP2 protein, which plays a vital role in activating the nonstructural protein complex by cleaving the proteins into subunits of nsP1, nsP2, nsP3 and nsP4. The model was further validated by molecular dynamics simulation and various validation

Table 4 Physiochemical descriptors calculated by Qikprop simulation

Lead molecules ^a	$QP \log P(o/w)$ ^b	$QP \log S$ ^c	$QP PCaco$ ^d	$QP \log HERG$ ^e	$QP PMDCK$ ^f	% Human oral absorption ^g
27943	-1.690	-1.582	0.345	0.090	0.269	1
21362	2.238	-4.109	112.378	-5.323	46.585	3
ASN 01107557	2.228	-5.106	54.703	-6.284	31.997	3
ASN 01541696	1.727	-4.691	95.051	-5.903	90.432	3

^a Ligand IDs: 43077: Binding Database; 21362: TOSLab; ASN 01107557 and ASN 01541696: Asinex database

^b $QP \log P$ for octanol/water (-2.0 , -6.5)

^c Predicted aqueous solubility, $\log S$. S in mol dm^{-3} is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid (-6.5 , -0.5)

^d Apparent Caco-2 permeability (nm/s) (<25 poor, >500 great)

^e $\log HERG$, HERG K⁺channel blockage (concern below -5)

^f Apparent MDCK permeability (nm/s) (<25 poor, >500 great)

^g % Human oral absorption in GI ($\pm 20\%$) (<25% is poor)

tools. Again, the model was subjected to flexible peptide docking and further e-pharmacophore mapping was carried out. Ligands that had a fitness score of more than 1.0 were subjected to a rigid docking study. As per our docking analysis, the residues Gln1039, Lys1045, Glu1157, Gly1176, His1222, Lys1239, Ser1293, Glu1296 and Met1297 show crucial interactions with the nonstructural protein complex to be cleaved, and were considered an individual functional unit. Chikungunya virus replication and propagation depends on the nsP2 protein; so a chemical compound that inhibits this protein by targeting the key residues specified above will be potentially applicable therapeutically. Based on the docking results, we can report four chemical compounds that may be potential inhibitors of nsP2 protease. Furthermore, the backbone structural scaffolds of these four lead compounds could serve as building blocks in the design of drug-like molecules for the treatment of Chikungunya viral fever. Besides targeting the Chikungunya virus, the inhibitors may act against other members of the *Alphavirus* genus due to the high sequence similarity among alphavirus proteins, which thus provides a clear potential path towards the identification of broad-spectrum drugs.

Acknowledgments The authors like to thank the Department of Bioinformatics, Alagappa University, Karaikudi, India for its support and providing the facilities for this work.

References

- Sourisseau M, Schilte C, Casartelli N, Trouillet C, Guivel-Benhassine F, Rudnicka D, Sol-Foulon N, Roux KL, Prevost M-C, Fsihi H, Frenkiel M-P, Blanchet F, Afonso PV, Ceccaldi P-E, Ozden S, Gessain A, Schuffenecker I, Verhasselt B, Zamborlini A, Saib A, Rey FA, Arenzana-Seisdedos F, Despres P, Michault A, Albert ML, Schwartz O (2007) Characterization of reemerging Chikungunya virus. *PLoS Pathog* 3:804–817
- Schuffenecker I, Itaman I, Michault A, Murri S, Frangeul L, Vaney M-C, Lavenir R, Pardigon N, Reynes J-M, Pettinelli F, Biscornet L, Diancourt L, Michel S, Duquerroy S, Guigon G, Frenkiel M-P, Brhin A-C, Cubito N, Despres P, Kunst F, Rey F, Zeller H, Brisse S (2006) Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *PLoS Med* 3:1058–1070
- Ng LFP, Chow A, Sun YJ, Kwek DJC, Lim PL, Dimatata F, Ng LC, Ooi EE, Choo KH, Her Z, Kourilsky P, Leo YS (2009) IL- β , IL-6, and RANTES as biomarkers of Chikungunya severity. *PLoS ONE* 4:e4261
- Vanlandingham DL, Hong C, Klingler K, Tsetsarkin K, McElroy KL, Powers AM, Lehane MJ, Higgs S (2005) Differential infectivities of O'nyong-nyong and Chikungunya virus isolates in anopheles gambiae and *Aedes aegypti* mosquitoes. *Am J Trop Med Hyg* 72:616–621
- Chopra A, Anuradha V, Lagoo-Joshi V, Kunjir V, Salvi S, Saluja M (2008) Chikungunya virus aches and pains: an emerging challenge. *Arthritis Rheum* 58:2921–2922
- Pardigon N (2009) The biology of chikungunya: a brief review of what we still do not know. *Pathol Biol* 57:127–132
- Takkinen K (1986) Complete nucleotide sequence of the nonstructural protein genes of Semliki Forest virus. *Nucleic Acids Res* 14:5667–5682
- Khan AH, Morita K, MdC P, Hasebe F, Mathenge EGM, Igarashi A (2002) Complete nucleotide sequence of Chikungunya virus and evidence for an internal polyadenylation site. *J Gen Virol* 83:3075–3084
- Strauss EGS, Strauss JH (1986) Structure and replication of the alphavirus genome. The *Togaviridae* and *Flaviridae*. Plenum, New York, pp 35–90
- Lulla A, Lulla V, Tints K, Ahola T, Merits A (2006) Molecular determinants of substrate specificity for Semliki Forest virus nonstructural protease. *J Virol* 80:5413–5422
- Perri S, Driver DA, Gardner JP, Sherrill S, Belli BA, Dubensky TW Jr, Polo JM (2000) Replicon vectors derived from Sindbis virus and Semliki Forest virus that establish persistent replication in host cells. *J Virol* 74:9802–9807
- Zhang D, Tözsér J, Waugh DS (2009) Molecular cloning, overproduction, purification and biochemical characterization of the p39 nsp2 protease domains encoded by three alphaviruses. *Protein Expr Purif* 64:89–97
- Vasiljeva L, Merits A, Auvinen P, Kaariainen L (2000) Identification of a novel function of the alphavirus capping apparatus. *J Biol Chem* 275:17281–17287
- Li L, Darden T, Hiskey R, Pedersen L (1996) Homology modeling and molecular dynamics simulations of the Gla domains of human coagulation factor IX and its G[12]A mutant. *J Phys Chem* 100:2475–2479
- Kevin JB, Edmond C, Huafeng X, Ron OD, Michael PE, Brent AG, John LK, Istvan K, Mark AM, Federico DS, John KS, Yibing S, David ES (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. In: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing. ACM, Tampa
- Shaw DE (2005) A fast, scalable method for the parallel evaluation of distance-limited pairwise particle interactions. *J Comput Chem* 26:1318–1328
- Bairoch A, Boeckmann B, Ferro S, Gasteiger E (2004) Swiss-Prot: juggling between evolution and stability. *Brief Bioinform* 5:39–55
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Russo AT, White MA, Watowich SJ (2006) The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease. *Structure* 14:1449–1458
- Schrödinger, LLC (2009) Prime, version 2.1. Schrödinger, LLC, New York
- Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118:11225–11236
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291
- Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins Struct Funct Bioinf* 17:355–362
- PyMOL (2010) PyMOL molecular graphics system website. <http://www.pymol.org>, accessed 2010
- Bowers KJ, Dror RO, Shaw DE (2006) The midpoint method for parallelization of particle simulations. *J Chem Phys* 124:184109–184111
- Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for

- proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105:6474–6487
27. Jørgensen AM, Topiol S (2008) Driving forces for ligand migration in the leucine transporter. *Chem Biol Drug Des* 72:265–272
 28. Lauria A, Ippolito M, Almerico AM (2009) Inside the Hsp90 inhibitors binding mode through induced fit docking. *J Mol Graph Model* 27:712–722
 29. Wang H, Aslanian R, Madison VS (2008) Induced-fit docking of mometasone furoate and further evidence for glucocorticoid receptor 17[alpha] pocket flexibility. *J Mol Graph* 27:512–521
 30. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47:1739–1749
 31. Salam NK, Nuti R, Sherman W (2009) Novel method for generating structure-based pharmacophores using energetic analysis. *J Chem Inf Model* 49:2356–2368
 32. Nabuurs SB, Wagener M, de Vlieg J (2007) A flexible approach to induced fit docking. *J Med Chem* 50:6507–6518
 33. Duffy EM, Jorgensen WL (2000) Prediction of properties from simulations: free energies of solvation in hexadecane, octanol, and water. *J Am Chem* 122:2878–2888
 34. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 23:3–25

Solvation counteracts coulombic repulsion in the binding of two cations to a model hexapeptide

Hongqi Ai · Chong Zhang · Wei He · Kwaichow Chan · Qiang Li

Received: 8 December 2010 / Accepted: 14 February 2011 / Published online: 29 March 2011
© Springer-Verlag 2011

Abstract In the present paper we employ a series of dicationized (K^+ & C^+ ; $C=H, Li, Na, \text{ or } K$) hexapeptide ($G_6C^+K^+_x$, $x=1, 2, \dots \text{ or } 5$) complexes in gas phase (GP) and aqueous phase (AP) as models to mimic the key characteristics of the real biosystems. An interesting phenomenon is observed that the binding properties of K^+_x to G_6C^+ species change in the two phases, i.e., those positive binding energies in the GP become negative ones in the AP. Then we probe the origin of property change of the binding energies in the two different phases and associate these changes with some biological phenomena.

Keywords Change of binding energy · $G_6C^+K^+_x$ complexes · Hydration · Origin

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1026-3) contains supplementary material, which is available to authorized users.

H. Ai (✉) · W. He
Shandong Provincial Key Laboratory of Fluorine Chemistry and Chemical Materials, School of Chemistry and Chemical Engineering, University of Jinan,
250022 Jinan, Shandong Province, People's Republic of China
e-mail: chm_aihq@ujn.edu.cn

C. Zhang
Department of Chemistry and Technology, Liaocheng University,
Liaocheng 252059, People's Republic of China

K. Chan
Department of Natural Science, Albany State University,
Albany, GA 31705, USA

Q. Li
School of Medicine and life, University of Jinan,
250022 Jinan, Shandong Province, People's Republic of China

Introduction

We present here a systematic study on the change origin of binding energy of a metal ion in the dicationic hexapeptide system to gain deeper insight into the origin of the driving force behind the commonly found biological systems, such as the KcsA selectivity filter of ion channel and proteins aggregation. The common feature of these biological systems is that two or more positive charged ions, which are interacting with each other in close proximity, are involved. The binding energy (ΔE_b) or binding free energy (ΔG) of the metal ion (M^+) in such a system is a positive value when the dicationic separation on host biomolecule is short range [1–6]:



Reaction (1) above indicates that energy will be released if the M^+ -(Reactant- C^+) bond of the product is broken, where C^+ = cation. Reaction (1) can also be described as a system containing any two like charged ions (anions or cations) with positive binding energy; such a system has been observed both experimentally [7] and by computer simulations [1–6]. Interestingly, the product in reaction (1) still stays in one of the minima on the potential energy surfaces (PES) [1–7]. The electrostatic repulsion of the two cations normally contributes the lion's share of the positive value in ΔE_b [2]. Obviously the repulsion is locally unfavorable for the stability of the system, but it might as well be the driving force [8] or a pre-requisite of biological engines [9].

The local repulsion is known to play a significant role for biomolecules in vivo. The well known example is the rapid transport of potassium ions through the KcsA selectivity filter within an ion channel [10]. Strong repulsion between the two K^+ ions binding to TVGYG

residues of the KcsA selectivity filter will result in the release of energy ΔE_b when either of two K^+ -O bonds along the peptide is broken. Structurally, the KcsA selectivity filter is a 12 Å short pore consisted of four juxtaposed chains of TVGYG residues, whose sequence is fairly conservative across a variety of biological species [10, 11], belonged to subunit of protein tetramer arranged in C_4 symmetry. Each of these TVGYG chains contains four carboxyl oxygens and one hydroxyl oxygen aligned longitudinally and sticking out toward to the center of the pore, forming five linear binding sites for trapping potassium ions or water molecules. As such, each of the binding sites rests in the center of eight surrounding carboxyl oxygens which occupy, in a sense, the corner vertices of a tetragon [9–11]. Only two to three K^+ ions can occupy the binding sites in the narrow short pore at a time. A potassium ion in the pore thus experiences a strong repulsion from another K^+ ion in the channel in manner analogous to K^+ ion experienced a strong push from the C^+ on our dicationic-hexapeptide models employed below, where repulsion can extend ~ 10 Å or more before structural changes occurred and negative binding energy settled in. Another example of dication separation is reported by Gross and Williams that a charge separation of 11.5 Å is found between the $H^+ \cdots X$ ($X=H, Li, Na, K$) in cyclic Gramicidin S (peptide + $H + X$) $^{2+}$ [12] in gas phase (GP). Gross and Williams also points out that the repulsion of H^+ and X^+ is the primary driving force behind the increase in proton transfer rates [13–18] and dissociation [19] observed in the double or multiply protonated ions in GP.

Besides the repulsion of cations bound to carbonyl oxygens and amino nitrogens [11, 20] on peptide chains of the protein, attentions have also been paid to the interaction between the multipositively charged ions of the peptide and its hydrophilic surrounding [21–30]. For example, a series of studies suggest that the interaction of positively charged guanidinium (Gdm^+) groups within and between protein subunits should be the crucial factor to alleviate or prevent the misfolding of proteins [25] because of their charge repulsion. In AP, however, Gdm^+ is weakly hydrated [26] and Gdm^+ ions tend to form like-charged contact ion pairs [27–29]. Moreover, the binding energy between two Gdm^+ ions becomes a negative value [30]. Obviously, it is solvent effect to change the property of the binding energy from positive to negative. Meanwhile, behaviors of multication interaction in an organism could be partially inferred, if not indirectly revealed, by investigating their hydrated cases because it is found that the maximum hydrated number of K^+ ion is seven to eight within the first solvation shell [31–33], similar to the coordination number of K^+ ion in the “KcsA selectivity filter” [34]. Multihydration of a metal cation implies that the metal cation can simultaneously coordinate with oxygens from several water molecules. Such a coordination can also be employed to simulate the interaction

between a metal ion and several carbonyl/hydroxyl oxygens, nitrogens or their hybrids in biomolecules [12].

Many important functions of biomolecules are too complicated to be studied quantum mechanically, yet without which the full understanding of some of the key functions cannot be achieved. To circumvent this problem, models that mimic the key characteristics of the real system are often employed. In this study, we employ dicationic hexapeptide $G_6C^+K_x^+$ and its hydrates as models to shed lights on the origin of property changes of binding energy ΔE_b of the $G_6C^+ \cdots K_x^+$ bond in GP and AP as well the transition from positive energy in GP to negative energy in AP, from which the properties of KcsA selectivity filter, Gramicidin S and protein aggregation processes are indicated and correlated. The dicationic hexapeptide system is small enough to be treated quantum mechanically and yet capture many of the key features without losing grip of reality when C and x are being adjusted. In the model $G_6C^+K_x^+$, G_6 represents a linear hexapeptide (hexaglycine), C^+ stands for a cation chosen from the list $\{C^+ = H^+, Li^+, Na^+, \text{ or } K^+\}$ that coordinates bidentately with both the terminal amino nitrogen and the carbonyl oxygen of the peptide G_6 , K_x^+ is the potassium cation bound to one of five possible sites denoted as x in the subscript, $x = 1, 2, \dots$ or 5 is the site number of one of the five carbonyl oxygens found on the G_6 chain. The exact order of the staggered oxygen sites is labeled according to Fig. 1. The bigger the x value, the further is the K^+ , bound to a carbonyl oxygen site, away from the C^+ along the linear peptide chain. The structure of $G_6C^+K_x^+$ changes accordingly as x increases; and so is the corresponding binding energy. Hence, the binding energy ΔE_b of a $G_6C^+K_x^+$ complex changes according to the type of cation C^+ bound at the terminal amino site and the dicationic separation r_{cx} , where r_{cx} is the distance between the terminal cation C^+ to the K^+ found at x th carbonyl oxygen. When r_{cx} is short range, i.e., when K^+ at a site x is near C^+ , the dicationic repulsion is strong, causing the binding energy ΔE_b to become positive and hence the local minimum structure of the hexapeptide unstable. When hydration effect is also added to the $G_6C^+K_x^+$ system and treated quantum mechanically, the relevance to the biological systems are found [9, 12, 30], to be discussed in later sections.

In a word, the change origin of binding energy in the models, $G_6C^+K_x^+$ and its hydrates, are discussed in detail in the present paper, the following categories are associated: (1) illustration of the importance of the involvement of short range repulsive dications as the likely key driving force for the rapid ion transport in the KcsA selectivity filter and for the increased rates of proton transfer [14–18] and dissociation [19] in the double or multiply protonated ions [35]; (2) insights into why the supporting tetrapeptides, which form the wall of most biological ion channel facilitating ions transport always in parallel to the long axis

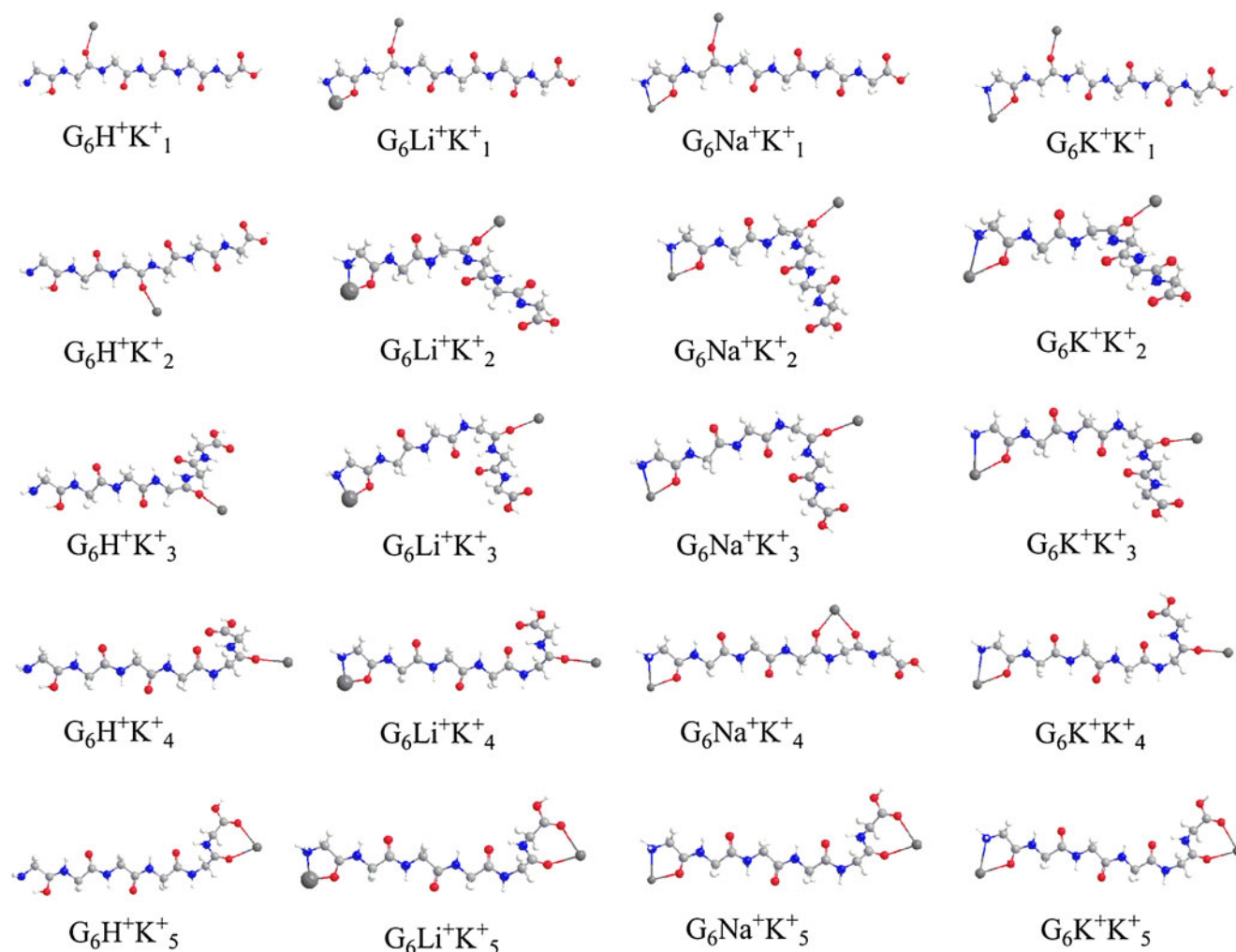


Fig 1 B3LYP/6-31 G*-optimized $G_6C^+K^+_x$ complexes (C=H, Li, Na, or K; $x=1,2,\dots, 5$)

of the peptides, is straight but not kinked ; (3) the rationality behind the distance between two neighboring cations in the selection filter is a short range and is typically shorter than 12 Å or else the linear peptide chain might kink beyond this range; (4) the origin of the cause of the change of binding energy of dicationic systems, such as Gdm^+ ion pairs, from positive in GP into negative in AP; (5) the effect of hydration on the shortrange of dicationic repulsion, structural stability and the transition of energy from positive GP to negative AP.

Computational details

Nomenclature and definitions used in the computation

The following definitions are provided to clarify some of nomenclature of the dicationic hexapeptide systems mentioned extensively in the text.

A *complex* refers to an instance or a special case of the $G_6C^+K^+_x$. For instance, $G_6Li^+K^+_1$ is a complex and $G_6K^+K^+_3$ is a different complex. A collection of complexes forms *isomers*, *families*, and *classes* according to definitions below. $G_6C^+K^+_x$ thus by itself stands for a complex that takes on any one of the possible ionic species and one of the five oxygen position x if the word isomer, family or class is not mentioned.

An *isomer* refers to two or more complexes of the same kind of C^+ but with different x values. When the C of $G_6C^+K^+_x$ assumes as a specific ion species, its two members of the family, for instance, $G_6Li^+K^+_1$ and $G_6Li^+K^+_4$, form isomers to each other; hence $G_6Li^+K^+_x$ constitutes five isomers.

A *family* refers to the collection of all possible isomers, e.g., $G_6Li^+K^+_1, G_6Li^+K^+_2, \dots, G_6Li^+K^+_5$ or $G_6Li^+K^+_x$ ($x = 1..5$) forms the family of $G_6Li^+K^+$. By definitions, $G_6Li^+K^+_x$ ($x = 1..5$) and $G_6Na^+K^+_x$ ($x = 1..5$) are two different families. For convenience, the term ‘family

$G_6Li^+K_x^+$ or ‘ $G_6Li^+K_x^+$ family’ means synonymously as $G_6Li^+K_x^+$ ($x = 1..5$).

An order n refers to complexes that have the same value x but C^+ s from all different family. For example, $G_6C^+K_2^+$ forms complexes of the order 2 which contains the members $\{G_6H^+K_2^+, G_6C^+Li_2^+, G_6Na^+K_2^+, \text{ and } G_6K^+K_2^+\}$. They are therefore a total 5 orders in our study: orders 1 $G_6C^+K_1^+$, ..., to order 5 $G_6C^+K_5^+$.

A Class refers to a collection of one or more families. For example, the four families $G_6H^+K_x^+$, $G_6Na^+K_x^+$, $G_6Li^+K_x^+$, $G_6K^+K_x^+$ ($x = 1..5$) form a class. In fact, $G_6C^+K_x^+$ ($C = H, Li, Na, \text{ or } K; x = 1..5$) represents the complete class of complexes involved in our calculation. A class can reduce to a family when the class has only one family as a member. For instance the subclass $\{G_6Na^+K_x^+, G_6Li^+K_x^+, G_6K^+K_x^+\}$ form an alkali metal class while $\{G_6H^+K_x^+\}$ forms a hydrogen subclass, which is also the hydrogen family.

$G_6C^+K_x^+$ ($C = Li, Na, \text{ or } K; x = 1..5$) is the alkali ion subclass.

$G_6C^+K_x^+(m \text{ or } -n)W$ where m or n is an integer that refers to a hydrated complex with m or n water molecules W bound to C_x^+ or K_x^+ . The geometries are optimized by the computation; for example, $G_6Na^+K_4^+4W$ means that the sodium ion is hydrated by four water molecules, whereas $G_6Na^+K_1^+-1W$ means that the K_1^+ is hydrated by one water molecule.

$G_6C^+K_x^+(m+n)W/F$ where m and n are integers, refers to as a hydrated complex with ‘ m ’ water molecules (W) bound to C_x^+ at the amino terminal and ‘ n ’ W or F molecules bound to K_x^+ at oxygen site x . F is formaldehyde; F/W means bound F or W at the K^+ .

$G_6K^+K_x^+(m+n)W1$ is an isomer of $G_6K^+K_x^+(m+n)W$. For instance, $G_6K^+K_1^+4+6W1$ is the isomer of $G_6K^+K_1^+4+6W$, with different spatial distribution of six waters at the K_1^+ site

$K_x^+ - nW$ refers to as the K^+ at site x is hydrated or bound by ‘ n ’ water molecules and n is an integer.

Selections of molecular geometry and computational method

We first optimized the stable linear G_6C^+ structures with C^+ bound at the amino terminal of the linear G_6 in GP. Note that the curved conformational structure can also be obtained if the starting geometry of the G_6 is not linear. Due to our focus on mimicking the key characteristics of some real biosystems such as KcsA selectivity filter, only the linear G_6C^+ structures are reserved to bind another K^+ at its x site of carbonyl oxygens in turn. The K^+ goes far away from the C^+ gradually along the peptide backbone just like a transfer within an ion pore. Moreover, the K^+ binding guarantees the linear array of the peptide chain

between C^+ and K^+ due to the electrostatic repulsion between two ions. Thus each optimized $G_6C^+K_x^+$ structure may be just a local minimum and does not always correspond to the global minimum which is not our profound concern. Our frequency calculations assert that all these structures do occupy minima (NIMAG = 0) on the potential energy surfaces. These are enough for our following discussion.

All GP geometry optimizations and vibrational frequency calculations are performed at the gradient-corrected DFT level using the 3-parameter fit of exchange and correlation functional of Becke (B3LYP) [36, 37] that includes the correlation functional of Lee, Yang, and Parr (LYP), as implemented in Gaussian 03 package of program Revision C.02 [38]. Basis set employed in these calculations is 6-31 G* [39].

The binding energy, including the correction of basis set superposition error (BSSE) [36], of a series of $G_6C^+K_x^+$ complexes is single-point calculated at the B3LYP/6-311++G**//6-31G* level according to the following equation (Eq. 2):

$$\Delta E_b = E_{G_6C+K_x} - (E_{G_6C^+} + E_{K_x^+}) \quad (2)$$

where $E_{G_6C+K_x}$, $E_{G_6C^+}$ and $E_{K_x^+}$ stand for the energy of the peptides $G_6C^+K_x^+$ and G_6C^+ in their optimal geometries and energy of a free K^+ , respectively. To probe AP binding energy, polarizable conductor calculation model (CPCM) [40] and solvent excluding surface are employed. Due to different effects of cavity models on the binding energy [41], six different atomic radii models, UAO, UAKS, UFF, UAHF, PAULING and BONDI are utilized to choose the optimal one that is in best agreement with the available theoretical or experimental data. To verify these results, MP2//6-311++G**//B3LYP/6-31G* calculations are also performed with the chosen atomic radii model. As a refinement to the implicit CPCM-calculation, the explicit hydration calculation is also performed for various combinations of $G_6C^+K_x^+$ complexes being considered. In detail, whenever C^+ bound at the amino terminal (N) in $G_6C^+K_x^+$ is saturated by m water molecules, its counterpart K^+ bound at one (x) of five carbonyl oxygens will also be surrounded by n water molecules. The binding energy of K^+ in reaction (1) now becomes that of its hydrates (K^+-nW , n is an integer and $W = \text{water}$) energy, i.e., the $E_{K_x^+}$ in Eq. 2 becomes the energy of the corresponding hydrated K^+ -clusters, which is obtained from the optimization of the dissociated K^+-nW species of corresponding $G_6C^+K_x^+$ hydrates.; $E_{G_6C^+}$ will also turn into the energy of hydrates $G_6C^+(m)W$ if C^+ is also hydrated by m water molecules.

Previous studies affirmed that the maximum coordination number of potassium ion in ion channel is eight [11, 34], but in water the maximum coordination number is found to be

seven, according to a recently updated study [31]. Thus the most reasonable scheme to probe the explicit hydration effect occurred at the second K ion site is to hydrate the C^+ site up to the seven water molecules first (plus a carboxyl oxygen of the peptide). In this study different structural arrangements for multihydration at the two cation sites (C^+ and K^+_x) have been systematically constructed to probe factors contributing to the optimal coordination number and way.

Results and discussion

Changes of binding energy from GP to implicit AP

The change of binding energy in relationship to the location of K^+ coordinated to carbonyl oxygen at site x ($x=1..5$ integer) is depicted in polynomial fits in Fig. 2. The plot reveals that the binding energy ΔE_b in GP is positive when $x \leq 3$ if the dications involved are alkali metals as in $G_6Li^+K^+_x$, $G_6Na^+K^+_x$ and $G_6K^+K^+_x$; but when hydrogen takes the place of C as in $G_6H^+K^+_x$, ΔE_b is positive when $x \leq 4$, which is in exact agreement with the previous prediction [5]. This effect of shifting x from site 3 to site 4 is due to protonation at the carbonyl oxygen of the amino terminal, which allows more charge density to disperse into the amino nitrogen than cationization of alkali metal ion at the same site [42], thus giving rise to the difference in the transition from positive to negative energy in ΔE_b . In general, the positive binding energy of the class $G_6C^+K^+_x$ ($x=1..5$) complexes decreases as the distance r_{cx} between dications C^+ and K^+ increases [5, 6]; which improves the dicationic peptide stability accordingly. For example, the ordering of binding energy of $G_6Na^+K^+_x$ family of complexes is $G_6Na^+K^+_5 > G_6Na^+K^+_4 > G_6Na^+K^+_3 > G_6Na^+K^+_2 > G_6Na^+K^+_1$ and hence the relative stability of this family of

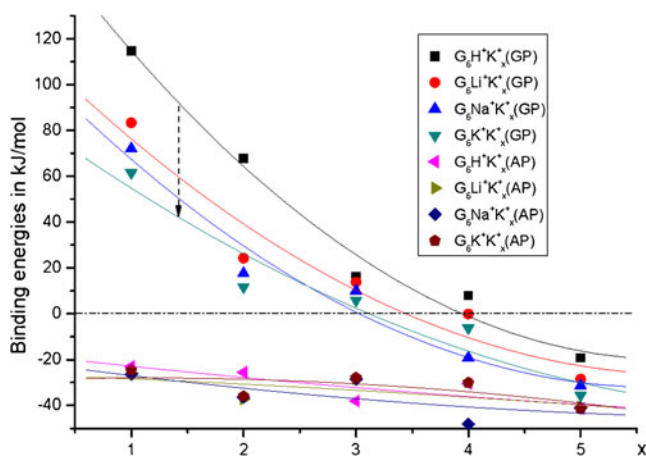


Fig 2 BSSE-corrected binding energies of series of $G_6C^+K^+_x$ complexes respective in the GP and AP

complexes. The site number of coordinated oxygen is therefore responsible for the critical change of ΔE_b from positive to negative is $x = 3$ and $x = 4$ for classes $G_6C^+K^+_x$ and $G_6H^+K^+_x$ respectively. Accordingly, the distances r_{cx} are 16.179, 16.541, and 16.787 Å for $G_6Li^+K^+_3$, $G_6Na^+K^+_3$ and $G_6K^+K^+_3$, respectively. The identical distance r_{cx} of $G_6H^+K^+_4$ is 19.110 Å, indicating the stronger repulsion between $H^+...K^+_x$ than that between $C^+...K^+_x$. These r_{cx} distances are listed in the Table S1, the supporting information (SI). In combination with the change of trend curve of binding energy in Fig. 2, we can estimate the turn point of binding energy change, i.e., from positive to negative for classes $G_6C^+K^+_x$ and $G_6H^+K^+_x$ is at about $r_{cx} = 17.0$ and 20.0 Å, respectively. Figure 1 clearly shows that the staggered peptide chain between the C^+ and K^+_x arranges in a line due to the effective repulsion between the two ions. It can be deduced that the peptide chain between the two ions will curve once the $r_{cx} > 17.0$ Å for the $G_6C^+K^+_x$ class and $r_{cx} > 20.0$ Å for the $G_6H^+K^+_x$ class due to the absence of electrostatic repulsion [43].

In sharp contrast to GP, in AP both the stability ordering and the ΔE_b 's of $G_6C^+K^+_x$ complexes change drastically. Detailed data of relative energy and binding free energy (ΔG) are listed in Tables 1 and 2 respectively. Note that binding energies in the AP presented in Table 2 are obtained by employing the PAULING radii. Pauling radii are preferred because our MP2/6-311++G** and B3LYP/6-311++G** calculations for K^+ hydration free energy in AP (see Table S2 in SI.) reveal that results generated by either PAULING or BONDI radii are more consistent with the available theoretical one [44] than those obtained by the other three radii models. The two radii chosen were once employed successfully for some similar systems [41]. Due to the lack of available theoretical and experimental data, the relative energy ordering in Table 2 calculated by B3LYP/6-311++G**//6-31G* are further cross-checked by the results of MP2/6-311++G** //B3LYP/6-31G*. Both methods offer the same rules for stability ordering although MP2 results have slightly larger than that of the B3LYP in absolute value. Only B3LYP results are used in the following discussions because of the saving in CPU time and computation accuracy equivalent to the MP2 method.

Table 1 reveals that the ordering of relative energy of these complexes changes somewhat in the AP relative to GP. Data in Table 2 lists all binding energies in AP that turn into negative values. Results show that the trend of change of these binding energies follows the stability ordering that has been established from ranking the relative energy of the corresponding hexapeptide family of complexes. This establishes that the ΔE_b is a major contributing factor to the stability. Most of the binding energy differences and the stability energy differences of

Table 1 Dipole moment (DM, Debye) in GP and relative energies (kJ mol⁻¹) in GP and AP (in parenthesis) of series of G₆C⁺K⁺_x

Complex	G ₆ H ⁺ K ⁺ ₅	G ₆ H ⁺ K ⁺ ₄	G ₆ H ⁺ K ⁺ ₃	G ₆ H ⁺ K ⁺ ₂	G ₆ H ⁺ K ⁺ ₁
ΔE	0.0(0.0/0.0)	25.5(11.3/15.5)	33.9(3.8/13.4)	84.0(16.3/29.3)	129.8(18.8/33.5)
DM	2.0	7.0	12.4	32.8	47.2
	G ₆ Li ⁺ K ⁺ ₅	G ₆ Li ⁺ K ⁺ ₄	G ₆ Li ⁺ K ⁺ ₃	G ₆ Li ⁺ K ⁺ ₂	G ₆ Li ⁺ K ⁺ ₁
ΔE	0.0(0.0/0.0)	26.8(-/-)	40.2(13.0/16.7)	51.5(4.2/13.8)	108.7(15.1/29.3)
DM	7.1	–	9.7	24.1	49.8
	G ₆ Na ⁺ K ⁺ ₅	G ₆ Na ⁺ K ⁺ ₄	G ₆ Na ⁺ K ⁺ ₃	G ₆ Na ⁺ K ⁺ ₂	G ₆ Na ⁺ K ⁺ ₁
ΔE	0.0(0.0/0.0)	10.9(-7.5/-1.7)	39.3(12.6/16.3)	47.7(4.2/13.8)	100.4(15.1/28.0)
DM	5.7	16.5	8.6	23.2	47.4
	G ₆ K ⁺ K ⁺ ₅	G ₆ K ⁺ K ⁺ ₄	G ₆ K ⁺ K ⁺ ₃	G ₆ K ⁺ K ⁺ ₂	G ₆ K ⁺ K ⁺ ₁
ΔE	0.0(0.0/0.0)	27.2(11.7/16.3)	38.9(14.2/18.0)	45.2(5.4/15.1)	94.1(17.6/31.4)
DM	4.5	9.5	8.3	21.6	44.9

Two values in parentheses refer to the results obtained, respectively, at the B3LYP and MP2 levels. Dipole moments are taken from the B3LYP level

the complexes in each family are less than 16.7 kJ mol⁻¹, indicating that aqueous solvent can effectively disperse the electrostatic repulsion from two positive charge centers (C⁺ and K⁺). Moreover, the regular change of both positive ΔE_b and relative stability of these complexes as a function of the distance r_{cx} between C⁺ and K⁺ binding sites, which is observed in the GP in Fig. 2 and Table 1, disappears.

Analysis on the binding energy changes from GP to implicit AP

What is the driving force causing positive binding energy in GP to become negative in AP? Obviously, solvent effect plays the key role reducing the electronic repulsion between two cation-centers [1, 8, 12–19] that keeps the binding energy in GP positive.

To show this effect intuitively, we perform bonding composition on these complexes in both GP and AP. The results are also displayed in Table 2. It is obvious that changes in binding free energy in AP (ΔG_a) are derived from the change of GP-binding free energy, the polarized solute-solvent interaction (ΔG_{a1}), and non-electrostatic terms (ΔG_{a2}) consisted of cavitation energy, dispersion energy and repulsion energy. A comparison clearly shows that the contribution of ΔG_{a2} is far less than the first two major terms. Due to minor BSSE value in B3LYP/6-311++G** calculations (<1.3 kJ mol⁻¹), ΔG_a values do not include the BSSEs. If we can treat GP-electronic binding energy (ΔE_g) in the same manner as GP-free binding energy, the sum of ΔE_g, ΔG_{a1} and ΔG_{a2} is almost equal to ΔG_a. By comparing the above three terms, we find that GP-binding energy is becoming more and more negative as the distance between the two positive charge centers increases, clearly confirming that gradual decrease in repulsion between the two ion centers leads to the

increase of K⁺_x in binding strength. In contrast, ΔG_{a1} not only increases gradually, but it also compensates effectively for the loss of binding strength in the GP and eventually keeps the negative free energy varying from a few to dozens of kJ mol⁻¹. The contribution of ΔG_{a2} does not follow the same rule as ΔG_{a1}. It is rather like a supplemental adjustment to the change of the entire binding free energy. For example, ΔG_{a2}'s of G₆H⁺K⁺₄ and G₆H⁺K⁺₃ are positive (2.5 kJ mol⁻¹), while their ΔG_{a1}'s are negative enough (-34.7/-49.0) to compensate for the ΔE_g's (8.0/16.3). For G₆H⁺K⁺₁ the non-electrostatic contribution (-25.1) is favorable for its binding strength and stability in AP although its ΔG_{a1} (-108.4) does not compensate for the loss of ΔE_g (114.6). These results reveal that stronger polarized solute-solvent interaction favors systems with shorter C⁺-K⁺ distance, and that the polarized solute-solvent interaction energy gradually decreases as the cationic radius (C⁺) increases.

Explicit hydration effect on the changes in binding energy

G₆K⁺K⁺₁ is selected to discuss the explicit hydration effect on the changes in binding energy as it has the strongest repulsive interaction between its two ions and thus is the most unstable structure among its G₆K⁺K⁺_x isomers. Relative to its isomers, G₆K⁺K⁺₁ presents the closest analogy in the study of repulsive interaction between ions to the KcsA selectivity filter. Calculations have therefore been extended to incorporate multi-hydrates of G₆K⁺K⁺₁ in order to elucidate the explicit hydration effect.

Selection of hydrated water number at the C⁺ and K⁺ sites

In Fig. 3, six hydrates of G₆K⁺K⁺₁ from mono, di, ... to hexahydrates, i.e., G₆K⁺K⁺₁W, G₆K⁺K⁺₁2W to G₆K⁺K⁺₁6W

Table 2 GP-deformation energies (ΔE_{d1}) of G_6C^+ species and BSSE-corrected binding energies (ΔE_g), and AP-binding energies (ΔE_a) and binding free energies (ΔG_a). Energy in kJ mol^{-1} and distance r in Å

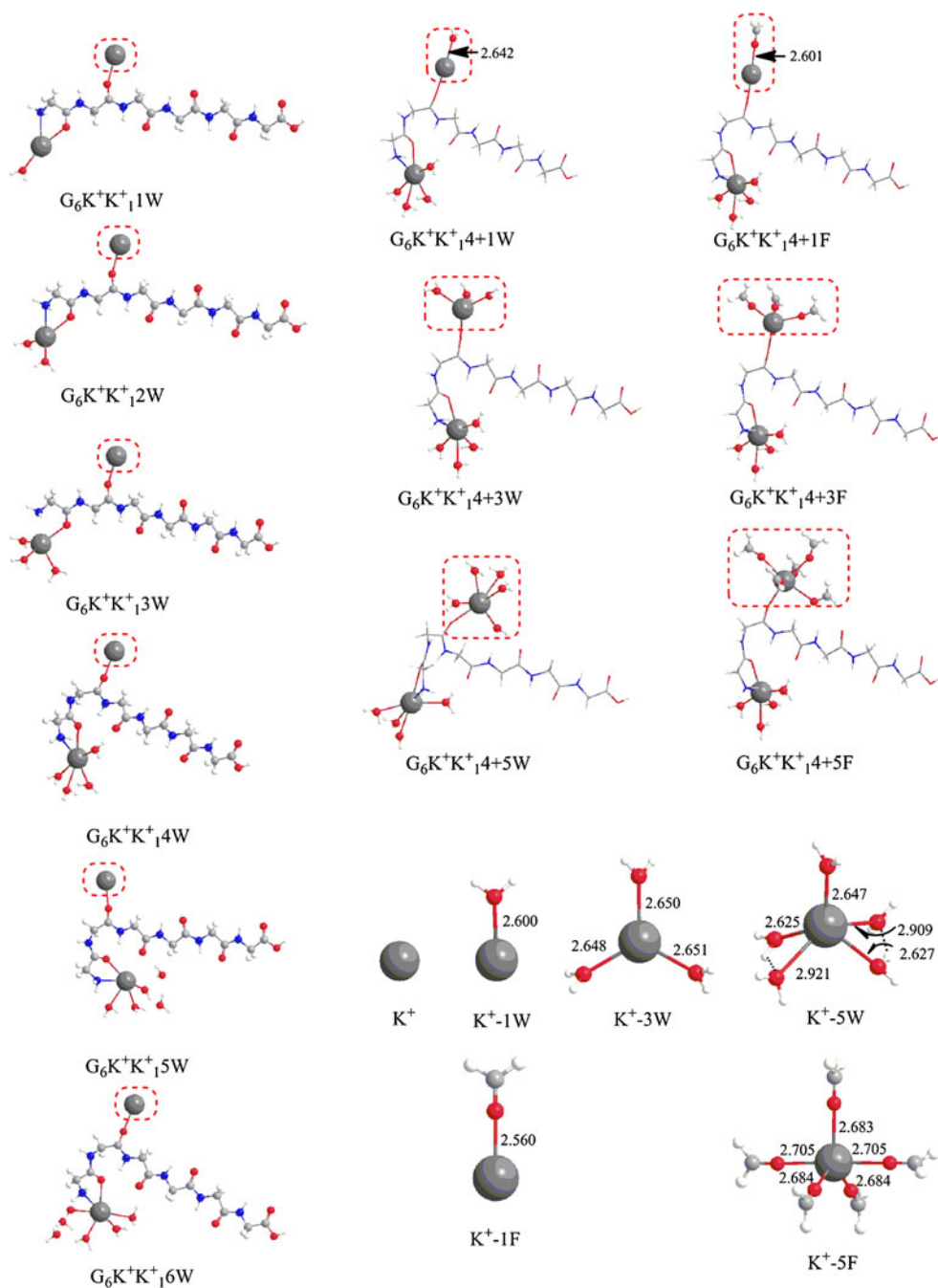
Complex	$G_6H^+K^+_1$	$G_6H^+K^+_2$	$G_6H^+K^+_3$	$G_6H^+K^+_4$	$G_6H^+K^+_5$
$\Delta E_g/\text{B3LYP}$	114.6	67.8	16.3	7.9	-19.2
$\Delta E_a(\text{B3LYP}/\text{MP2})$	-23.0/-27.2	-25.5/-31.8	-38.1/-47.3	-30.5/-45.2	-41.4/-60.2
$\Delta G_a/\text{B3LYP}$	-23.8	-31.8	-32.2	-25.1	-46.0
$\Delta G_{a1}/\text{B3LYP}$	-108.4	-92.5	-49.0	-34.7	-10.0
$\Delta G_{a2}/\text{B3LYP}$	-25.1	-1.7	2.5	2.5	-7.5
ΔE_{d1}	13.4	19.7	8.0	11.3	26.8
$r/\text{Å}$	2.617	2.532	2.482	2.486	2.579(r_1)/2.631(r_2)
	$G_6Li^+K^+_1$	$G_6Li^+K^+_2$	$G_6Li^+K^+_3$	$G_6Li^+K^+_4$	$G_6Li^+K^+_5$
ΔE_g	83.3	24.3	13.8	-0.1	-28.5
$\Delta E_a(\text{B3LYP}/\text{MP2})$	-25.9/-31.0	-37.2/-46.4	-28.5/-43.5	-/-	-40.6/-59.8
$\Delta G_a/\text{B3LYP}$	-25.9	-31.4	-21.3	-	-44.8
$\Delta G_{a1}/\text{B3LYP}$	-104.6	-53.1	-37.7	-	-0.2
$\Delta G_{a2}/\text{B3LYP}$	-0.4	5.4	6.7	-	-4.2
ΔE_{d1}	14.6	7.9	9.2	10.9	26.4
$r/\text{Å}$	2.565	2.484	2.480	2.482	2.576(r_1)/2.630(r_2)
	$G_6Na^+K^+_1$	$G_6Na^+K^+_2$	$G_6Na^+K^+_3$	$G_6Na^+K^+_4$	$G_6Na^+K^+_5$
$\Delta E_g/\text{B3LYP}$	72.0	17.6	10.0	-19.2	-31.4
$\Delta E_a(\text{B3LYP}/\text{MP2})$	-26.4/-32.6	-36.4/-46.9	-28.5/-44.4	-48.1/-61.9	-41.0/-60.2
$\Delta G_a/\text{B3LYP}$	-25.9	-30.5	-22.2	-51.9	-44.8
$\Delta G_{a1}/\text{B3LYP}$	-93.7	-47.7	-35.6	-20.1	2.1
$\Delta G_{a2}/\text{B3LYP}$	0.4	6.3	6.7	-3.8	-4.2
ΔE_{d1}	14.2	7.9	12.6	44.8	24.3
$r/\text{Å}$	2.551	2.480	2.478	2.623(r_1)/2.552(r_2)	2.574(r_1)/2.630(r_2)
	$G_6K^+K^+_1$	$G_6K^+K^+_2$	$G_6K^+K^+_3$	$G_6K^+K^+_4$	$G_6K^+K^+_5$
$\Delta E_g/\text{B3LYP}$	61.5	11.7	5.9	-6.3	-35.6
$\Delta E_a(\text{B3LYP}/\text{MP2})$	-24.7/-29.7	-36.0/-46.0	-28.0/-42.7	-30.1/-44.8	-41.4/-60.3
$\Delta G_a/\text{B3LYP}$	-23.4	-31.0	-21.3	-24.7	-45.6
$\Delta G_{a1}/\text{B3LYP}$	-86.6	-44.8	-33.5	-23.8	1.3
$\Delta G_{a2}/\text{B3LYP}$	1.3	5.9	6.7	5.4	-4.2
ΔE_{d1}	14.6	8.4	9.6	11.3	26.8
$r/\text{Å}$	2.539	2.477	2.477	2.479	2.572(r_1)/2.630(r_2)

ΔG_{a1} and ΔG_{a2} denote the polarized solute-solvent interaction contribution and non-electrostatic term contributions on the ΔG_a of complexes in AP, respectively. r refers to the distance between the second K^+ and its binding site of peptide

are displayed. Also displayed in Fig. 3 are three multihydrates of $G_6K^+K^+_14W$, $G_6K^+K^+_14+1W/F$, $G_6K^+K^+_14+3W/F$, and $G_6K^+K^+_14+5W/F$. Their dissociated water- K^+ clusters are also included in the figure. It is found that only $G_6K^+K^+_14W$ but not any other members of the hydrated $G_6K^+K^+_1mW$ ($m \neq 4$, $m = 1, 2, \dots$) is a suitable model to further hydrate/formylate at the K^+_1 site. And to differentiate it from K^+_1 , K^+ at the amino-terminal is denoted as $K^+_{(N)}$. First, the binding between the K^+_1 and its carboxyl oxygen at site x in general gets stronger as the number of hydrated waters at the $K^+_{(N)}$ site increases. Table S3 shows such a trend of binding energy from that of the $G_6K^+K^+_11W$, $G_6K^+K^+_12W$, ..., to $G_6K^+K^+_15W$. The binding energy of $G_6K^+K^+_14W$ ($\Delta E_{ba} = 19.7 \text{ kJ mol}^{-1}$) turns out to be the

least among the five hydrates, thus representing the strongest binding of $O \dots K^+_1$ bond [33]. It is also found that the four water molecules in $G_6K^+K^+_14W$ coordinate to the $K^+_{(N)}$ directly; whereas two of the water molecules in $G_6K^+K^+_15W$ and $G_6K^+K^+_16W$ are bound to the second hydration shell due to the doping coordination of three peptide groups and the limitation of steric effect. For these reasons $G_6K^+K^+_14W$ should be employed as an optimal model to discuss the further hydration at the K^+_1 site for the sake of saving CPU time and attaining reasonable computational accuracy. Due to our focus on hydration at the K^+_1 site, the optimal radii $CN = 8$ [11, 34] should be chosen. Here we mainly choose the lower coordination number ($CN = 5, 6, 7$) to discuss in the following sections

Fig 3 Amino terminal K-hydrated complexes ($G_6K^+K^+_1(1\sim 6)W$) and further carboxyl K-hydrated $G_6K^+K^+_14+$ (1,3,5)W/F as well as one of their corresponding dissociation products (K^+-nW/F , $n=0\sim 5$). Distance in angstrom



because we find that in most cases the maximum CN within the first hydration shell for K^+_1 is 6 (five water molecules plus a carboxyl oxygen, see Fig. S1) due to the steric effect of the peptide chain of G_6C^+ , and additional water molecules have to be attached the second shell.

Binding energy changes and its contributions for these complexes of $G_6C^+K^+_x(m+n)W$

It is obvious that the binding energy ΔE_b in formula (2) becomes here the energy difference of $G_6C^+K^+_x(m+n)W/F$ and its two optimized species G_6C^+mW peptide and K^+-

nW/F cluster. Note that K^+-nW/F cluster derives from $G_6C^+K^+_x(m+n)W/F$ and thus its relaxed structure in the calculation of binding energy does not always correspond to the global minimum but maybe a local minimum. Anyway, it presents a real hydrated K^+ cluster which involves the formation of the $G_6C^+K^+_x(m+n)W$ and the energy difference of the “relax” cluster and “rigid” one in $G_6C^+K^+_x(m+n)W$ configuration stands for the deformation energy, which contributes to the binding energy negatively. Table S4 shows that despite the difference in hydration number and manner, the positive binding energy of the $G_6K^+K^+_1$ has never changed into negative because of the

hydration [1–6, 9]. In contrast, the positive binding energy of $G_6K^+K_1^+$ is far larger than that of $G_6K^+K_1^+4+1W$, $G_6K^+K_1^+4+3W$, and $G_6K^+K_1^+4+5W$, but only somewhat larger than that of $G_6K^+K_1^+4W$. This result confirms that hydration at the $C_{(N)}^+$ site of peptide chain can strengthen the binding between K_1^+ and the O site; however, hydration at the K_1^+ site will weaken it.

Multihydration at the K_1^+ does not necessarily lead to a decrease of the binding strength of all complexes. The irregularity of the changes of binding energy depends on the hydrated number and the way hydration is performed. Hydration at the K_1^+ from 1 W, 3 W, 5 W to 9 W leads to less change in the binding energy but more deformation of the structure. For example, the deformation energy of these hydrates induced by hydration is 6.7, 4.6, 24.7 and 86.6 kJ mol⁻¹ (notice the sudden jump) respectively. Among these values, contribution from K_1^+-nW ($n=1, 3, 5, \text{ and } 9$) species are 0.8, 1.3, 16.3 and 64.9 kJ mol⁻¹ respectively; and from the G_6C^+ species the contribution becomes 5.9, 3.3, 8.4, 21.8 kJ mol⁻¹ respectively. Thus, for the complexes with less hydration at the K_1^+ site, the deformation of G_6C^+ dominates; whereas for those with more hydration at the K_1^+ site, deformation of K_1^+-nW dominates. As the number of hydrated water increases at K_1^+ , the deformation extends progressively and yet gradually the bonding distance of $G_6C^+...K_1^+$ from 2.539→2.604→2.823→2.829 Å, thereby weakens the binding strength.

Since the difference between the vertical binding energy (ΔE_{bv}) and the corresponding adiabatic binding energy (ΔE_{ba}) is the deformation energy, the detailed changes in vertical and adiabatic binding energy as a function of the number of bound water molecules is depicted in Fig. S2. Results show that the values of ΔE_{bv} of $G_6K^+K_1^+4+5W1$, $G_6K^+K_1^+4+6W1$, and $G_6K^+K_1^+4+7W1$ are negative, whereas their counterparts of $G_6K^+K_1^+4+5W$, $G_6K^+K_1^+4+6W$, and $G_6K^+K_1^+4+7W$ are all positive values, indicating that the former is more stable than the latter. A comparison of relative energy clearly shows that $G_6K^+K_1^+4+5/6W1$ are indeed more stable than their counterparts $G_6K^+K_1^+4+5/6W$. However, $G_6K^+K_1^+4+7W1$ is an exception, less stable than the $G_6K^+K_1^+4+7W$ although the later has a relatively more compact structure than the former. The structural distinction between the two types of multihydrates ($G_6K^+K_1^+4+nW1$ vs $G_6K^+K_1^+4+nW$, $n = 5, 6, 7$) lies only in the change in spatial distribution of water molecules bound to the K_1^+ (Fig. S1). Thus what governs their relative energy is not in agreement with what governs their binding energy in such multihydrated system. This phenomenon implies: the same number of hydrated water but different manner of hydration will result in great difference in binding strength of the metal ion; and a more stable structure does not necessarily imply a stronger $G_6C^+...K^+$ binding. This important result provides us with a deep

insight as into why a specific *steric configuration* is necessary for functional biomolecules such as the ion channel. Obviously, such a pore configuration meets the ion transfer with less binding strength and stronger driving force [9].

The hydration effect on the turning point of binding energy

The binding energies of isomers $G_6K^+K_3^+$ and $G_6K^+K_4^+$ are 5.9 and -6.3 kJ mol⁻¹ respectively, indicating that a qualitative transition (turning point) of the binding energy occurred somewhere between $3 < x < 4$. Will such a qualitative change also occur as a result of hydration? Two particular hexapeptide complexes are chosen as models for hydrating their K_x^+ sites, for the sake of saving CPU times, in order to obtain their respective pentahydrated ~ enneahydrated configurations. These hydrates are exemplary because they cover the maximum coordination number of the K^+ in both AP [31–33] as well as in the ion channel [34]. Their corresponding geometries and binding energy are displayed in the Fig. S3 and Table S4 respectively. Both ΔE_{ba} and ΔE_{bv} of the $G_6K^+K_4^+(5\sim 8)W$ complexes become positive, suggesting that hydration at the carbonyl-terminal K_4^+ indeed weakens the binding of $G_6K^+...K_4^+mW$. Likewise, ΔE_{ba} of $G_6K^+K_3^+(5\sim 7)W$ complexes are also positive and greater than their isomeric $G_6K^+K_4^+(5\sim 7)W$ counterparts, indicating that the turning point of binding energy disappears and that the shorter the distance between two ions, the stronger is their repulsion, and such a trend cannot be changed by the hydration.

Our analysis also shows that deformation energy (ΔE_d) of the $G_6K^+K_3^+(5\sim 7)W$ complexes are far greater than that of the corresponding isomeric $G_6K^+K_4^+(5\sim 7)W$; and that the deformation energy derives mainly from the relaxation of dissociated K^+-nW clusters (i.e., ΔE_{d1}). That the deformation energy of $G_6K^+K_3^+mW$ is greater than that of $G_6K^+K_4^+mW$ is closely related to the $K^+...K^+$ distance, which is shorter in the former than the latter, and hence greater repulsion. The result affirms directly why two K^+ ions in the ion channel prefer to shorter but limited separate distance [10]: one is to ensure rapid transfer rate of K^+ ions, another is to keep the dynamic stability of the ion channel.

Comparisons of hydration and formulation at the K_1^+ sites

Because the behavior of the carboxyl oxygens in biological molecules is more similar to the oxygen of formaldehyde (F) than to the oxygen of water molecules, the water bound at the K_1^+ in $G_6K^+K_1^+4+1/3/5W$ are replaced by formaldehyde's to form complexes $G_6K^+K_1^+4+1/3/5F$ to simulate dicationic hydration in biological environment. The

slightly higher adiabatic binding energy ΔE_{ba} of $G_6K^+K^+_{14+3/5F}$ than that of the corresponding $G_6K^+K^+_{14+3/5W}$ in Table S4 indicates that the formaldehyde molecule has stronger binding affinity to the K^+ than water molecule and as a result, weakens the strength of $G_6K^+ \dots K^+_{14+3/5F}$ bond. The trend is that the energy gap between the corresponding complexes with and without formaldehyde increases as the number of formaldehyde coordination increases. This result further confirms that multi-coordinated architecture favors the transfer of K^+ ion due to the larger and positive $G_6K^+ \dots K^+_1$ binding strength. The binding energy gap between $G_6K^+K^+_{14+3/5F}$ and $G_6K^+K^+_{14+3/5W}$ also indicates the feasibility for the K^+ transfer from the water cavity into the channel [11], and even is thus likely the key energy source for driving K^+ transfer from the water cavity to the ion channel [33, 44].

Effect of metal ions on the binding energy in hydration

The binding strength of K^+_x varies according to the types of metal ions (C^+) actually bound to the amino terminal. Table 2 documents this effect. How about the effect in hydration? Monohydrates $G_6C^+K^+_x-1W$ are taken as examples to discuss this issue for brevity's sake (figures are not shown). Table S3 lists their binding energy. When $C^+ = H^+$, monohydration at the K^+_x increases the binding strength of $G_6H^+K^+_5$ only, but decreases the rest of the family members $G_6H^+K^+_x$ ($x = 1, 2, 3, \text{ or } 4$). When $C^+ = Li^+$, monohydration increases the binding strength of isomers $G_6Li^+K^+_2$, $G_6Li^+K^+_3$ and $G_6Li^+K^+_5$. Similar trend is also observed for $C^+ = Na^+$ or K^+ . Thus there is a difference between H^+ and alkali metal ions in ligand- K^+_x binding; and monohydration at the K^+_x site will, consequently, lead to changes in binding strength. For the class of monohydrated $G_6C^+K^+_4-1W$ ($C' = Li, Na, \text{ or } K$) complexes, their binding energy decreases abnormally in comparison to their non-hydrated counterparts. Additionally, only the binding energy of double potassium involved complex $G_6K^+K^+_4-1W$ (-3.3 kJ mol^{-1}) has ever become negative after it is monohydrated. Generally, monohydration at the K^+_x drives the already positive binding energy of those complexes even higher, i.e., further decreases the binding strength of K^+_x -peptide; conversely, monohydration renders those complexes with negative binding energy even more negative, i.e., further strengthens the binding energy of those already negative in energy. Similar trend or rule is also observed in the $G_6K^+K^+_x-1F$ complexes. From the entire class ($C^+ = H^+, Li^+, Na^+, \text{ to } K^+$), the binding energy of the second K^+_x in the corresponding $G_6C^+K^+_x-1W/F$ complexes increases gradually, revealing why rate of proton transfer from their $(\text{peptide}+H+M)^{2+}$ to dipropylamine decreases as the radius of alkali metal ion increases [12].

Origin of different ΔE_b signs and corresponding biological relevance

Why does PCM approach yield negative binding energies while explicit solvent modeling in most cases led to positive values? Which is more relevant to physical reality? Obviously, both cases are relevant to the physical reality in different status.

The positive value of ΔE_b obtained in these explicit hydrates derives from that the number of water molecules and their arrangement only around the cation just can decrease peptide-ion interactions but are not sufficient to effectively screen the cation-cation repulsion. The phenomenon is more like that in some specific biological structures, such as KcsA K^+ channel [9, 10], where positive ΔE_b [9, 10] and that only dihydrated case for K^+ is optimal [35] were reported. It implies and just as discussed above that the number of hydrated molecules, specific configuration and structural constraints to coordinate K^+ ions respond to specific demands. In GP or non-hydrated cases, the positive ΔE_b becomes larger and indicates a stronger repulsion between two cations, which results in a portion of the peptide chain tethered between the two, the cations prefers to be stretched as shown in Fig. 1 and observed by Williams et al in their experiment [17] and theoretical simulation [12]. In a word, the magnitude and sign of ΔE_b in GP or explicit hydrates depend on the type of cations chosen for C^+ , the distance between the two cation centers, and the solvated number and hydrated manners surrounding the two cations. These results further highlight the importance of discussion on different hydration of those with multication involved systems

On the other hand, if the cation-cation repulsion is screened completely by solution, as that in an implicit CPCM-solvent, then the negative binding energy will be obtained. The results can explain well why the guanidine group of arginine-arginine pairing via positively charged guanidinium (Gdm^+) groups can not exist in GP except within AP or between protein subunits [21–24]. Obviously, the binding energy between a pair of Gdm^+ in the latter environment becomes negative due to sufficient shield effect of hydration [30].

Conclusions

In this study, we employ dicationic hexapeptide $G_6C^+K^+_x$ ($C=H, Li, Na, K; x=1..5$) and its hydrates as models to investigate the change origin in binding energy of the $G_6C^+ \dots K^+_x$ bond from GP (positive) to AP (negative), from which the properties of KcsA selectivity filter, Gramicidin S and protein aggregation are indicated and correlated. The following significant conclusions can be drawn.

Our study of hexapeptide $G_6C^+K^+_x$ shows the positive binding energy induced by short range dicationic repulsion depends strongly on separation distance between the two ions, C^+ and K^+_x , which leads to change of the binding energy of K^+_x ion and the hexapeptide and causes the characteristic transition from positive to negative energy when the $C^+ \cdots K^+_x$ distance is more than 17.0 Å. As long as it is within this narrow range of distance in GP, the positive in energy can be maintained. Multihydration will not change the fate of the positive energy, but it will render the positive more or less in magnitude than null hydration in the following manners: those with higher positive value will become less positive while those with lower positive value will become more positive after multihydration. This explains why the pore length of selectivity center is limited (~ 12 Å) [10] and why the charge separated distance in the (peptide+H+X)²⁺ ions (X=H, Li, Na, K) gramicidin S is 11.5 Å and the peptide chain between two ions is extended or linear but not curved [12].

Explicit mono~multi-hydration will greatly affect the K-O(peptide) interaction but keep the binding energy of the K^+_x -O bonds always positive and the peptide chain between C^+ and K^+_x always extended and linear. It is because of that all these hydrated water molecules only locate around the K^+ and are not sufficient to effectively screen the cation-cation repulsion, and thus lead to a positive ΔE_b . This implies that the $C^+ \cdots K^+$ repulsion is still palpable in such a short range of peptide chain despite the two ions are multi-hydrated or multi-carboxylated. Hydration in different steric manners even with the same hydration number will yield completely different results in binding energy, which strongly suggests that specific steric structure is a pre-requisite for specific biological functions [12–19]. The implication of this realization helps shed light on the understanding of the structural function of an ion channel. Within an ion channel, the short range ion-ion repulsion is the driving force for the rapid ion transfer through the channel. It also keeps the peptide straight, allowing for straight tetrapeptides to form as a supporting wall for most biological ion channel, thereby accommodating the ions to be transported in parallel to the long axis of the peptides. The reason these peptide chains do not kink is also due to each ion in the channel is eight-fold coordinated with carboxyl/hydroxyl oxygens from TVGYG residues [4, 11]. In the absence of such a perfect configuration to accommodate the ions inside the ion channel, kinks of peptides around the ions will happen. Therefore keeping the four-fold pentapeptides, which form the wall of the pore, straight is a pre-requisite for the stability of the ion channel. Hence these two apparently very different characteristics (fast transfer rate and short, straight channel comprised of tetrapeptides of length~

12 Å) are really two sides of the same coin [1–7]. In fact they are in dynamic equilibrium. Additionally, the difference between the binding energy of hydrated complexes and its formulated (carboxylated) counterparts also predicts the feasibility of K^+ ions entering the pore of an ion channel from the water cavity [11].

In an implicit CPCM-solvent, the cation-cation repulsion is screened completely by solution, as a result, turning the positive binding energy into negative. This result clearly accounts for why in GP Gdm^+ ion pairs cannot exit (mutual repulsion) whereas in AP can (i.e., tend to form like-charged contact ion pairs [27–29], and their binding energy becomes negative [30]). Further analysis shows that stronger polarized solute-solvent interaction in the implicit AP favors systems with shorter C^+ - K^+ distance, and their interaction energy gradually decreases as the C^+ cationic radius increases. This accounts for why all individual ions involved in the complexes $G_6C^+K^+_x$, $x=1..5$ in AP have more or less the same negative binding energy although the positive binding energy in GP decreases along with the increase in cationic radius.

The change of positive ΔE_a in GP along with cationic radius is also closely related to the following facts: (1) The smaller the ion C^+ radius is, the more charge transfers to the peptide chain [12, 42]; hence the stronger repulsion between two cationic centers will become [8]. (2) The result is in good agreement with experiments of Williams et al. that rates of proton transfer decreases as ion radii (from Li^+ , Na^+ , to K^+) increase in the H^+ & M^+ involving peptides [12]. (3) Since K^+ has the largest ion radius and therefore the most positive binding energy among all four different types of C^+ ions. This fact dictates that $K^+ \cdots K^+$ pairs are the most preferred ions to be transported inside the ion channel. The radius of K^+ ion is the most suitable for spatial structure inside the pore for precision contact with the surrounding eight-fold coordinated carboxyl oxygens, which, in turn, keep optimal transfer rate possible.

Despite all the encouraging signs that the simple dicationic hexapeptide systems can be used as models to generate deeper insights to the understanding of some of the key characteristics and functions of KcsA selectivity filter of ion channel, Gramicidin S, and local protein-protein interaction, there are unavoidable shortcomings. First of all, our models are rather simple and are still quite different structurally from the real ion channel or the cyclic Gramicidin S molecule. The relevance of conclusions obtained in the present study is not yet one to one correspondence to direct observation and measurement obtained for the real biological systems. Thus that next work will be done is to employ a real protein molecule, such as an ion channel protein, to perform molecular dynamic simulation on whole molecule and its surroundings and then quantum calculation on the key functional fragment

of the molecule, in which the findings on the change of the binding energy of ions and ion-ion repulsion interaction reported here can be verified and extended.

Acknowledgments H. Ai gratefully acknowledges the support of National Natural Science Foundation of China (NSFC) (Grant Nos. 20973084 and 20573047), and Natural Science Foundation (NSF) (Grant No. Y2008B56) and Doctoral Fund (Grant No. 2007BS02009) of Shandong Province.

References

1. Ai H, Bu Y, Han K (2002) *J Chem Phys* 117:7593–7602
2. Ai H, Bu Y, Chen Z (2003) *J Chem Phys* 118:1761–1772
3. Ai H, Bu Y (2004) *J Chem Phys* 120:2208–2214
4. Ai H, Zhang C, Li Y, Zhang L, Li F (2007) *J Phys Chem B* 111:13786–13796
5. Ai H, Bu Y, Li P, Yan S (2005) *J Chem Phys* 123:134307(1–7)
6. Ai H, Li Y, Zhang C, Feng J (2007) *Chem Phys* 334:64–76
7. Kass SR (2005) *J Am Chem Soc* 127:13098–13099
8. Honig B, Nicholls A (1995) *Science* 268:1144–1149
9. Guidoni L, Torre V, Carloni P (2000) *FEBS Lett* 477:37–42
10. Doyle DA, Cabral JM, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, Chait BT, MacKinnon R (1998) *Science* 280:69–77
11. Ban F, Kusalik P, Weaver DF (2004) *J Am Chem Soc* 126:4711–4716
12. Gross DS, Williams ER (1996) *J Am Chem Soc* 118:202–204
13. Schnier PD, Gross DS, Williams ER (1995) *J Am Soc Mass Spectrom* 6:1086–1097
14. McLuckey SA, van Berkel GJ, Glish GL (1990) *J Am Chem Soc* 112:5668–5670
15. Gross DS, Williams ER (1995) *J Am Chem Soc* 117:883–890
16. Gross DS, Rodriguez-Cruz SE, Bock S, Williams ER (1995) *J Chem Phys* 99:4034–4038
17. Schnier PD, Gross DS, Williams ER (1995) *J Am Chem Soc* 117:6747–6757
18. Gross DS, Schnier PD, Rodriguez-Cruz SE, Fagerquist CK, Williams ER (1996) *Proc Natl Acad Sci USA* 93:3143–3148
19. Loo JA, Edmonds CG, Udseth HR, Smith RD (1990) *Anal Chim Acta* 241:167–173
20. Tissandier MD, Cowen KA, Feng WY, Gundlach E, Cohen MH, Earhart AD, Tuttle TR Jr, Coe JV (1998) *J Phys Chem A* 102:7787–7794
21. Auffinger P, Bielecki L, Westhof E (2004) *Structure* 12:379–388
22. Ledvina PS, Yao N, Choudhary A, Quioco FA (1996) *Proc Natl Acad Sci USA* 93:6786–6791
23. Vyas NK, Vyas MN, Quioco FA (2003) *Structure* 11:765–774
24. Pednekar D, Tendulkar A, Durani S (2009) *Proteins* 74:155–163
25. Marti DN, Bosshard HR (2004) *Biochemistry* 43:12436–12447
26. Mason PE, Neilson GW, Dempsey CE, Barnes AC, Cruickshank JM (2003) *Proc Natl Acad Sci USA* 100:4557–4561
27. Boudon S, Wipff G, Maigret B (1990) *J Phys Chem* 94:6056–6061
28. No KT, Nam KY, Scheraga HA (1997) *J Am Chem Soc* 119:12917–12922
29. Mason PE, Neilson GW, Enderby JE, Saboungi ML, Dempsey CE, MacKerell AD, Brady JW (2004) *J Am Chem Soc* 126:11462–11470
30. Vondrášek J, Mason PE, Heyda J, Collins KD, Jungwirth P (2009) *J Phys Chem B* 113:9041–9045
31. Jensen KP, Jorgensen WL (2006) *J Chem Theory Comput* 2:1499–1509
32. Marcus Y (1988) *Chem Rev* 88:1475–1498
33. Varma S, Rempe SB (2007) *Biophys J* 93:1093–1099
34. Zhou YF, Morais-Cabral JH, Kaufman A, MacKinnon R (2001) *Nature* 414:43–48
35. Aqvist J, Luzhkov V (2000) *Nature* 404:881–884
36. Becke AD (1993) *J Chem Phys* 98:5648–5652
37. Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37:785–789
38. Frisch MJ, Trucks GW, Schlegel HB (2004) *Gaussian 03, Revision C.02*. Gaussian Inc, Wallingford, CT
39. Boys SF, Bernardi F (1970) *Mol Phys* 19:553–566
40. Barone V, Cossi M (1998) *J Phys Chem A* 102:1995–2001
41. Wang Y (2008) *J Phys Chem C* 112:14297–14305
42. Rodriguez CF, Cunje A, Shoeih T, Chu IK, Hopkinson AC (2000) Michael Siu KW. *J Phys Chem A* 104:5023–5028
43. Wyttenbach T, Bushnell JE, Bowers MT (1998) *J Am Chem Soc* 120:5098–5103
44. Varma S, Rempe SB (2008) *J Am Chem Soc* 130:15405–15419

Application of electron conformational–genetic algorithm approach to 1,4-dihydropyridines as calcium channel antagonists: pharmacophore identification and bioactivity prediction

Nazmiye Geçen · Emin Sarıpınar · Ersin Yanmaz · Kader Şahin

Received: 14 December 2010 / Accepted: 16 February 2011 / Published online: 31 March 2011
© Springer-Verlag 2011

Abstract Two different approaches, namely the electron conformational and genetic algorithm methods (EC-GA), were combined to identify a pharmacophore group and to predict the antagonist activity of 1,4-dihydropyridines (known calcium channel antagonists) from molecular structure descriptors. To identify the pharmacophore, electron conformational matrices of congruity (ECMC)—which include atomic charges as diagonal elements and bond orders and interatomic distances as off-diagonal elements—were arranged for all compounds. The ECMC of the compound with the highest activity was chosen as a template and compared with the ECMCs of other compounds within given tolerances to reveal the electron conformational submatrix of activity (ECSA) that refers to the pharmacophore. The genetic algorithm was employed to search for the best subset of parameter combinations that contributes the most to activity. Applying the model with

the optimum 10 parameters to training (50 compounds) and test (22 compounds) sets gave satisfactory results ($R^2_{training} = 0.848$, $R^2_{test} = 0.904$, with a cross-validated $q^2 = 0.780$).

Keywords Dihydropyridines · Drug design · Electron conformational–genetic algorithm · Pharmacophore · 4D-QSAR

Introduction

In drug discovery and development, the design of novel drugs is an interdisciplinary, laborious, time-consuming and extremely expensive process guided by trial and error [1]. Therefore, the early prediction of the activity-related properties of drug candidates is an important problem in this area. A convenient and efficient method to predict biological activity from structural information is thus indispensable. QSAR studies express the biological activities of compounds as a function of their various structural features and describe how variations in biological activity depend on changes in chemical structure [2].

Dihydropyridines (DHPs) have a broad range of pharmacological actions as agents in bronchodilation, hepatoprotection and geroprotection, and as antiatherosclerosis, antidiabetes, antitumor, antimutagenic, antioxidant, anticonvulsant and antiradical agents [3]. 1,4-DHPs are also one of the largest and most studied groups of drugs for the treatment of cardiovascular disease, as they are a class of organic calcium channel blockers and clinically significant antihypertensive drugs [4–8].

Several recent QSAR studies on DHP derivatives containing substituted nitroimidazolyl, phenylimidazolyl,

N. Geçen
Department of Chemistry, Science and Art Faculty,
Siirt University,
Siirt 56100, Turkey

E. Sarıpınar (✉)
Department of Chemistry, Science Faculty, Erciyes University,
38039 Kayseri, Turkey
e-mail: emin@erciyes.edu.tr

E. Yanmaz
Department of Chemistry, Altınoluk Vocational College,
Balıkesir University,
Balıkesir, Turkey

K. Şahin
Faculty of Art and Science, Bitlis Eren University,
Bitlis, Turkey

and methylsulfonyl-imidazolyl at the C-4 position and different alkoxy groups at the C-3 and C-5 positions of the DHP ring have been reported using the hyperbox approach [9], genetic algorithm–partial least squares (GA-PLS) and principal component–genetic algorithm–partial least squares (PC-GA-PLS) [10], least squares support vector machines [11], neural networks (NN) and principal component analysis (PCA) [12]. CoMFA (comparative molecular field analysis), CoMSIA (comparative molecular similarity indices analysis), and GRID/GOLPE (generation of optimal linear PLS estimations) methods were applied by Schleifer et al. to a common pharmacophore model of calcium antagonistically active 1,4-DHP in order to find the correlation between interaction fields and biological activity [13]. Safarpour et al. [14] employed genetic algorithm–multiple linear regression (GA-MLR) and genetic algorithm–artificial neural network (GA-ANN) methods to a series of DHP derivatives containing substituted nitro-imidazolyl groups in order to elucidate the interrelation between physicochemical parameters and biological activity. In Safarpour et al.'s paper, they used a genetic algorithm (GA) as a variable selection tool; two different models were obtained by applying GA-MLR and GA-ANN. Artificial neural networks (ANN) have the characteristic of being suitable for processing data in which the relationship between the cause and its result cannot be exactly defined. It is widely applied in QSAR nonlinear modeling and has achieved substantial positive results. One advantage of neural networks is that they are naturally capable of modeling nonlinear systems. Disadvantages include a tendency to overfit the data, and a significant degree of difficulty in ascertaining which descriptors are most significant in the resulting model. In addition, one particular problem associated with most ANN models is their tendency to become overtrained (i.e., the compounds used for model development are perfectly described but the models lack the ability to predict new data). As a result, when faced with a test set of new observations, the predictive ability of such a model will be very poor. Another drawback of NN is the matter of interpretability. An important limitation of neural networks is their lack of ability to explain. Also, for complex tasks, it may be time consuming to train a neural network, as there are too many parameters to tune to achieve optimal performance [15]. MLR is widely used in QSAR modeling because of its simplicity [16, 17]. A drawback of multiple linear regression (MLR) is that it lacks the ability to handle complex nonlinear problems in its standard form, although nonlinear extensions are available to this end. In addition, the main drawback of MLR in the field of QSAR modeling is that it cannot be used when the number of explanatory variables exceeds the number of observations.

Because chemical interactions are three-dimensional events, QSARs often depend on the 3D molecular models

adopted for the chemicals under study. Although many biological properties are dependent on the 3D conformers of a molecule, in 3D-QSAR analysis each ligand molecule is represented by a single bioactive conformation and orientation. The other conformers are not analyzed or modeled in 3D-QSAR techniques. Current 3D-QSAR methods have three property sampling limitations: conformation, alignment and pharmacophore are usually restricted or fixed. Moreover, each of these three properties is held fixed when building a particular 3D-QSAR model. Generally, the conformation that is assumed to be the lowest energy conformation of the molecule is used to build the model [18].

To get an even better understanding of the relationship between biological activity and the structure of compounds, higher dimensional QSAR techniques can be used, but they are computationally expensive. 4D-QSAR methodology is an extension of the 3D-QSAR methodology developed by Hopfinger et al. which considers conformational information as the fourth dimension [19]. The fourth dimension refers to the possibility of representing each ligand molecule as an ensemble of conformations, thereby reducing the bias in identifying bioactive conformations and orientations. Conformation constraints and molecular flexibility strongly affect the bioactivity of flexible molecules [20].

The concept of the pharmacophore is widely used in drug design, and it is generally defined as a three-dimensional map of biological properties common to all active conformations of a set of ligands which exhibit a particular activity [21]. Conceptually, a pharmacophore is a distillation of the functional attributes of ligands that accomplish a specific task. Knowledge of the pharmacophore, or the 3D arrangement of features in the biologically active molecule that is responsible for its pharmacological activity, can help in the research and design of new or better drugs acting upon the same or related target. Many methods have been developed for pharmacophore detection and its applications [22]. The electron topological (ET) method proposed by Dimoglo et al. [23–25] and the electron conformational (EC) method [26–29] are some of the QSAR methods used to identify a pharmacophore. Additionally, the EC method, which was proposed by Bersuker et al., is used for bioactivity prediction based on a nonlinear mathematical model [30].

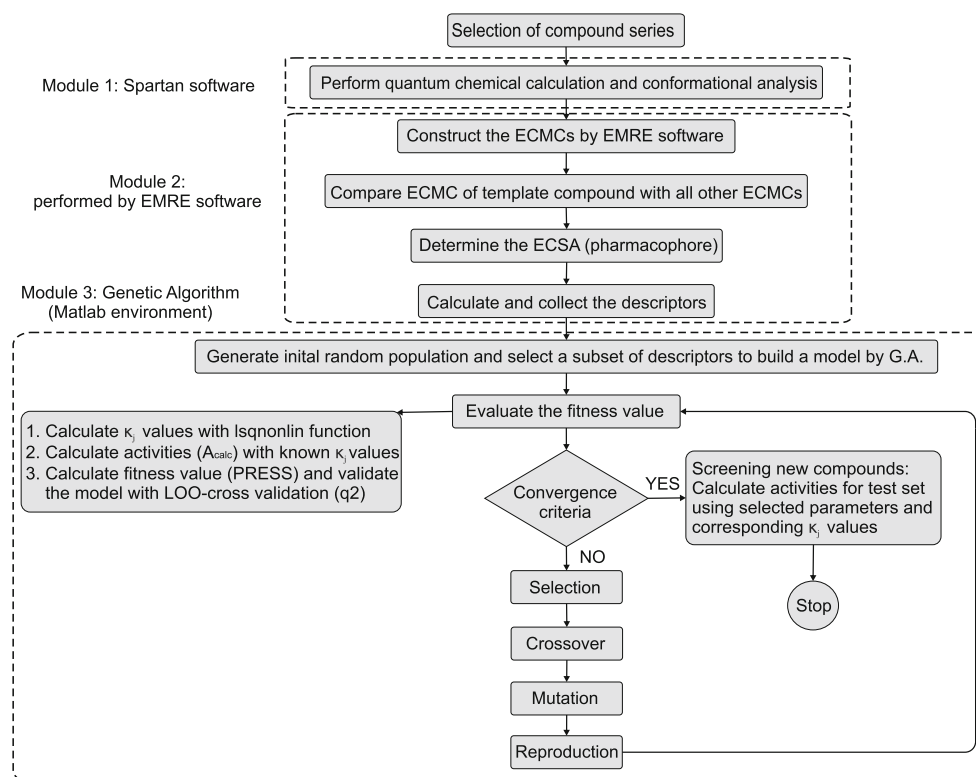
The physicochemical and structural properties of molecules are usually represented by a set of variables, and the assumption is that the molecule's activity is in some way related to the values of these variables. Not all of the descriptors generated for a specific molecule are significant in QSAR models. In order to select valuable descriptors and to build predictive models, various methods—such as principal component analysis (PCA) [31], partial least squares (PLS) [32], GA [33–38] and neural networks (NNs) [39]—have been widely used in many QSAR studies and methods.

In the present paper, the electron conformational–genetic algorithm (EC-GA) method is presented as a novel hybrid 4D-QSAR approach for pharmacophore identification and bioactivity prediction using the best subset of parameters selected by GA. This method is based on the generation of a conformational ensemble profile for each compound instead of only one conformation. In our previous studies, this method was successfully performed as a 4D-QSAR procedure to identify the pharmacophore and to predict the quantitative bioactivities of triaminotriazine derivatives [40] and benzotriazines as Src inhibitors, and to quantitatively predict activity [41] and penicillin derivatives [42].

Methods

The recently developed EC-GA method, which is described in more detail elsewhere [40, 41], consists of three main modules and is summarized in Fig. 1. The first module in the figure involves quantum-mechanical calculations for and conformational analysis of the compound series with Spartan software. The second module performed by the EMRE software we devised constructs and compares ECMCs, identifies the pharmacophore, and calculates (or collects) descriptors. The last module combines a GA and a cross-validation technique to calculate the bioactivities of training and test compounds, and to produce an EC-GA model that is based on the best selected subset of parameters.

Fig. 1 Flow chart depicting the proposed hybrid EC-GA method which is a combination of electron conformational and genetic algorithm methods. Three major modules are shown. Module 1: quantum mechanical calculation and conformational analysis by Spartan software; module 2: construction and comparison of ECMCs by EMRE software; module 3: calculation of bioactivity for training and test compounds, and construction of the EC-GA model using GA and a cross-validation technique (Matlab environment)



The structural and experimental data for the DHPs used in this study were taken from the literature [43–47] and are listed in Table 1. The number of conformers and the predictive activities of compounds corresponding to experimental values are also given in Table 1. The reported IC_{50} values were converted into corresponding pIC_{50} values. n indicates the number of CH_2 groups in the chain of the ester group. Symmetrical compounds are indicated by an asterisk in the table. Underlined numbers correspond to the 22 compounds of the test set, randomly selected from the original 72 compounds.

The three-dimensional structures of the DHPs were constructed and their geometries optimized by means of the semiempirical PM3 method using Spartan '06 v.1.2.0 software [48]. These energy-minimized structures were then subjected to conformational analysis using the Monte Carlo randomized search method over all rotatable bonds. For any given molecule, conformers with the lowest energy will have the largest weight in the ensemble of energetically accessible conformers. After performing a conformational analysis for each molecule, conformers with energies above $1.5 \text{ kcal mol}^{-1}$ of the global minimum were eliminated [49]. The remaining conformers were aligned to the lowest-energy conformer by overlapping their atomic positions. Conformers with different spatial orientations were kept. An individual data file was created for each conformer of the molecule. Each heavily populated conformer of each molecule with n atoms was described by its electron

Table 1 Molecular structures, (A^{exp}) and calculated (A^{calc}) activity values for the 1,4-DHP data set

(symmetrical) (unsymmetrical)

Compound ^[a]	R ¹	R ²	n ^[b]	NC ^[c]	pIC ₅₀	
					A ^{exp[d]}	A ^{calc[e]}
<u>1</u>	n-Butyl	Ethyl	0	4	11.4900	11.6329
<u>2</u> *	n-Butyl	n-Butyl	0	3	11.2900	11.9708
<u>3</u>	OCH ₃	i-Propyl	2	6	12.2100	12.6749
<u>4</u>	OH	Methyl	2	31	11.8900	12.0716
<u>5</u>	COCH ₃	Methyl	3	11	11.8300	11.0872
<u>6</u>	CN	i-Propyl	2	10	11.8300	9.8077
<u>7</u>	OH	i-Propyl	2	10	11.5300	11.1309
<u>8</u>	OH	Ethyl	2	22	11.4500	11.5609
<u>9</u>	OCH ₃	Methyl	2	17	11.3800	11.7633
<u>10</u>	OCH ₃	Ethyl	2	9	11.1300	11.5592
<u>11</u>	c-Hexyl	Methyl	0	7	9.0100	9.7298
<u>12</u>	c-Hexyl	Ethyl	0	5	8.4000	8.4260
<u>13</u>	c-Hexyl	Methyl	1	29	8.5700	8.8856
<u>14</u>	c-Hexyl	Ethyl	1	12	8.2600	8.4778
<u>15</u>	c-Hexyl	Methyl	2	16	8.4000	8.7287
<u>16</u>	c-Hexyl	Ethyl	2	3	7.8200	7.7390
<u>17</u>	c-Hexyl	Methyl	3	7	8.3300	8.1730
<u>18</u>	c-Hexyl	Ethyl	3	6	8.2800	8.2614
<u>19</u>	c-Hexyl	Methyl	4	8	8.3000	8.1383
<u>20</u>	c-Hexyl	Ethyl	4	2	8.2200	6.9898
<u>21</u>	c-Pentyl	Methyl	3	6	8.7300	8.8650
<u>22</u>	c-Pentyl	Ethyl	3	4	8.4000	8.9706
<u>23</u>	c-Propyl	Methyl	1	7	7.4800	8.5159
<u>24</u>	c-Propyl	Ethyl	1	12	7.1300	8.4972
<u>25</u>	Phenyl	Methyl	1	36	7.7000	8.0273
<u>26</u>	Phenyl	Ethyl	1	27	7.7000	8.3326
<u>27</u>	Phenyl	Methyl	2	17	7.7000	8.3310
<u>28</u>	Phenyl	Ethyl	2	3	7.5600	6.3912

Table 1 (continued)

29	p-Tolyl	Methyl	2	20	7.9000	7.5955
30	p-Tolyl	Ethyl	2	9	7.3200	7.1148
31	Phenyl	Methyl	3	5	7.6100	8.0397
32	Phenyl	Ethyl	3	12	7.4200	8.1316
33	Phenyl	Methyl	4	6	7.5700	7.5906
34	Phenyl	Ethyl	4	15	7.2100	7.6193
35	Phenyl	Methyl	5	7	7.1400	7.2063
36	Phenyl	Ethyl	5	7	7.0800	6.8748
37*	c-Hexyl	c-Hexyl	0	3	7.8500	8.0595
38*	c-Hexyl	c-Hexyl	1	6	9.3100	8.4140
39*	c-Hexyl	c-Hexyl	2	10	8.8800	8.7044
40*	c-Hexyl	c-Hexyl	3	2	7.5500	7.6168
41*	n-Propyl	n-Propyl	0	5	10.050	9.8523
42*	c-Propyl	c-Propyl	1	7	7.5700	7.9322
43*	c-Pentyl	c-Pentyl	3	3	7.8900	8.0784
44*	Phenyl	Phenyl	1	8	8.5800	7.7383
45*	Phenyl	Phenyl	2	11	8.4200	7.5843
46*	p-Tolyl	p-Tolyl	2	3	7.2800	7.9166
47*	Phenyl	Phenyl	3	4	7.3500	6.9486
48*	Phenyl	Phenyl	4	6	6.8900	6.9299
49*	Phenyl	Phenyl	5	2	5.9500	5.9936
50*	i-Butyl	i-Butyl	0	10	11.3600	10.4563
51	ONO ₂	Methyl	2	8	11.5900	11.9223
52	ONO ₂	Ethyl	2	13	10.9300	11.9010
53	ONO ₂	i-Propyl	2	21	12.0200	12.3812
54	ONO ₂	Methyl	3	20	11.3300	11.7322
55	ONO ₂	Ethyl	3	18	11.1100	12.2964
56	ONO ₂	i-Propyl	3	14	11.7500	12.2961
57	ONO ₂	Methyl	4	27	11.6000	10.6555
58	ONO ₂	Ethyl	4	11	10.8700	11.3638
59	CH(CH ₂ ONO ₂) ₂	Methyl	0	18	11.4100	11.6134
60	CH(CH ₂ ONO ₂) ₂	Ethyl	0	10	11.6600	11.8422
61	CH(CH ₂ ONO ₂) ₂	i-Propyl	0	8	11.6100	11.5918
62	N(CH ₃) ₂	Methyl	3	16	8.1400	8.9264
63	N(CH ₃) ₂	Ethyl	3	9	8.9600	8.8772
64	N(CH ₃) ₂	i-Propyl	3	15	9.3300	9.5562
65	N(CH ₃) ₂	Methyl	2	12	8.5000	9.1861
66	N(CH ₃) ₂	Ethyl	2	4	9.3100	9.4393
67	N(CH ₃) ₂	i-Propyl	2	8	10.1300	8.9605

Table 1 (continued)

68	t-Butyl	Methyl	0	24	9.7400	8.1090
69	t-Butyl	Ethyl	0	17	10.3100	8.8397
70	n-Pentyl	Ethyl	0	4	12.1300	10.0534
71	Methyl	Ethyl	0	15	11.1300	11.2662
72	n-Pentyl	Methyl	0	10	12.3300	12.3300

^aUnderlined numbers correspond to the 22 compounds of the test set, randomly selected from the original 72 compounds; symmetrical compounds are indicated by an asterisk

^bnumber of CH₂ chain in ester groups

^c number of conformers for each compound

^d experimental activity values of 1,4-DHP data set

^e calculated activity values of 1,4-DHP data set

conformational matrix of congruity (ECMC), where n varies from one molecule to another. Figure 2 illustrates an example of the ECMC calculated for the lowest-energy conformer of compound 72. While the diagonal elements of this matrix refer to the Mulliken charge of the corresponding atom in the molecule, the off-diagonal elements are either bond orders for chemically bonded pairs of atoms or interatomic distances for all other pairs of atoms. In this way, each ECMC contains both electronic and geometric characteristics, which are deemed to adequately represent the properties of the compounds under consideration [50].

In the next step, we chose the ECMC of the lowest energy conformer of the compound with the highest activity as a

template and compared it with the ECMCs of the lowest-energy conformers of the other compounds within given tolerances in order to reveal those matrix elements that are present in active compounds but absent from inactive or low-activity compounds [28]. In this study, 794 ECMCs (each of which corresponds to one conformer) obtained from the 1,4-dihydropyridine derivatives (72 compounds) were constructed and compared using EMRE software.

This comparison procedure resulted in several submatrices called electron conformational submatrices of activity (ECSA). In general, the tolerance values of high-activity compounds are lower than those of low-activity compounds. By gradually changing the limits of tolerance for

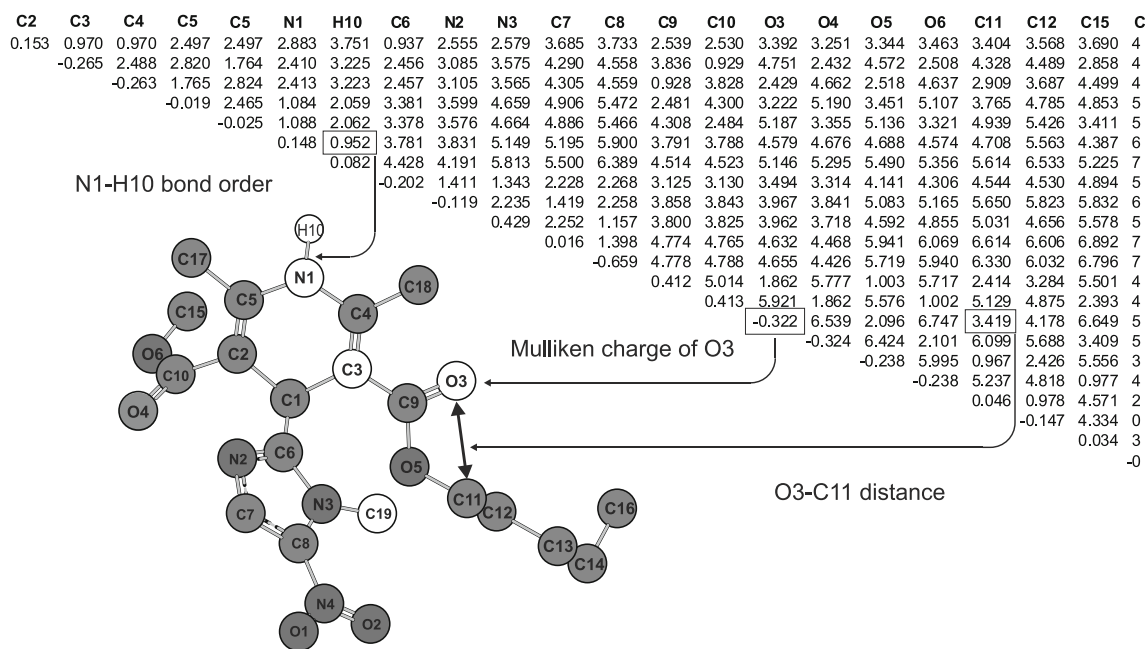


Fig. 2 ECMC of the lowest-energy conformer of the template compound (compound 72) with the highest activity in the series of 1,4-DHP derivatives. The diagonal elements refer to the Mulliken charges, while the off-diagonal elements are bond orders for

chemically bonded pairs of atoms and interatomic distances for nonbonded pairs. Hydrogen atoms bonded to carbon atoms are excluded from the ECMC for simplicity

diagonal and nondiagonal matrix elements, we finally obtained the tolerance limits of the ECSA that give the best separation of highly active compounds from similar inactive ones [49, 51]. In order to determine whether the resulting ECSA (pharmacophore) is good or bad, two probabilistic estimations from the literature were used. They were given as follows [25, 52]:

$$P_\alpha = (n_1 + 1)/(n_1 + n_3 + 2), \quad (1)$$

$$\alpha_a = (n_1 \times n_4 - n_2 \times n_3)/(m_1 \times m_2 \times m_3 \times m_4)1/2, \quad (2)$$

where n_1 and n_2 are the numbers of molecules in this class of compounds that are and are not active, respectively; n_3 and n_4 have analogous meanings for weakly active compounds; m_1 and m_2 are the numbers of molecules in the classes of active and weakly active compounds, respectively; $m_3 = n_1 + n_3$; $m_4 = n_2 + n_4$. While P_α (Eq. 1) relates only to active molecules, α_a (Eq. 2) relates to both active and weakly active compounds. Then, without setting any constraints on the tolerance values, the maximum tolerance values of the best-matching ECSA were calculated for each conformer of each compound. The submatrix (pharmacophore) resulting from this approach represents the 3D arrangements of the structural or chemical features of a drug. The elucidated pharmacophore for 1,4-DHP is given in the “Results and discussion” section (Table 2).

After determining the pharmacophore, the next step was to perceive and identify all features that may contribute to the biological activity. For this purpose, the contributions of the pharmacophore parameters and the parameterized out-of-pharmacophore groups to the biological activity were described by the following equation for the S function, which represents the sum of these effects [51]:

$$S_{ni} = \sum_{j=1}^N K_j a_{ni}^{(j)} \quad (3)$$

where κ_j are variational coefficients, $a_{ni}^{(j)}$ are parameters describing the j -th kind of property in the i -th conformation of the n -th compound, and N is the number of selected parameters. The relative weights of different parameters with respect to the biological activity of a compound are described by the constants κ_j .

Within the ensemble of conformers, the contribution of an individual entity to the total energy is determined by a normalized Boltzmann. Using the function S and taking into account the Boltzmann population of each conformation as a function of its energy and temperature, a formula for the activity is obtained that contains a sum of terms for all conformations. Considering some well-known general features of the substrate–enzyme (S–E) interaction, Bersuker et

al. presented a formula [49] that defines the nonlinear relationship between bioactivity and the parameters. This formula is given below:

$$A_n = A_l \frac{\sum_{i=1}^{m_l} e^{-\frac{E_{li}}{RT}} \sum_{j=1}^{m_n} \delta_{ni} [Pha] e^{-S_{ni}} e^{-\frac{E_{nj}}{RT}}}{\sum_{i=1}^{m_n} e^{-\frac{E_{ni}}{RT}} \sum_{j=1}^{m_l} \delta_{li} [Pha] e^{-S_{li}} e^{-\frac{E_{lj}}{RT}}} \quad (4)$$

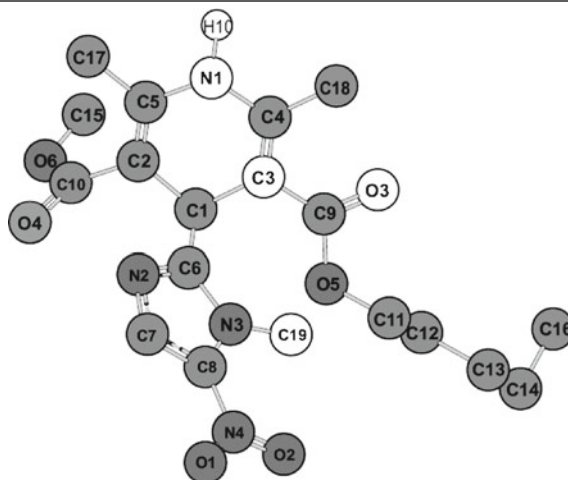
where δ is the Kronecker delta. This is a function of two variables that is 1 if the pharmacophore is present and 0 otherwise. A_n and A_l are the activities of the n -th compound and the reference molecule, respectively. m_n is the number of conformations of the n -th molecule. E_{ni} is the relative energy of the i -th conformation of the reference compound, E_{ni} is the relative energy of the i -th conformation of the n -th compound (in kcal mol⁻¹), R (kcal mol⁻¹ K⁻¹) is the gas constant, and T is the temperature in K.

The most popular algorithm for estimating the weights (κ_j) of the parameters is the nonlinear least-squares method, which considers that the best-fitting straight line is the line that minimizes the square of the error between the predicted activity, A_n^{calc} , and the observed activity, A_n^{exp} . We obtained the adjustable constants κ_j by performing a least-squares minimization operation on the function $\sum_n |A_n^{calc} - A_n^{exp}|^2$. With the κ_j constants determined in this way, we were able to estimate the expected activity of any molecular system using Eq. 4. This complex multidimensional data was investigated in the Matlab environment using a lsqnonlin function based on a general nonlinear least-squares fitting algorithm. The numbers κ_j , $j=1, 2, \dots, N$, obtained in this way characterize the weights of each of the $a_{ni}^{(j)}$ parameters in the overall APS/AG influence [49].

The active conformer is often not the lowest-energy conformer, and as a result the flexibility of compounds is inevitably overlooked [20]. Since relatively small energy differences between conformers can result in significant variations in electronic structure, in the 4D-QSAR [19] approach it is essential to account for the Boltzmann populations and the dynamics of these conformational substates of all compounds in order to understand the effects of all energetically stable conformers on the biological activity [53]. In this study, not only the lowest-energy conformation but also all reasonable conformers given in Table 1 entered the final activity formula (Eq. 4) in order to predict the bioactivity. To reduce the (lengthy) computation time associated with the number of conformers, a parallel GA—which is very effective at handling computationally expensive problems—was used to solve this problem and to find the best solution. The parameters of the conformers of each compound were carefully prepared using the EMRE program.

Table 2 ECSA (pharmacophore) of template compound (compound 72) for 1,4-dihydropyridine derivatives (a), tolerance matrix of ECSA for 38 compounds with high activity (b), tolerance matrix of ECSA

for 34 compounds with low activity (c), and tolerance values for all conformers (794). Pharmacophore atoms are shown in white



a) ECSA (pharmacophore) for template compound (compound 72)

C3	N1	H10	O3	C19	Pha Atoms
-0.263	2.413	3.223	2.429	4.126	C3
	+0.148	0.952	4.579	5.964	N1
		+0.082	5.146	6.760	H10
			-0.322	4.366	O3
				-0.163	C19

b) Tolerance matrix of ECSA for 38 compounds with high activity

C3	N1	H10	O3	C19	
±0.036	±0.023	±0.035	±0.016	±0.150	C3
	±0.045	±0.007	±0.161	±0.164	N1
		±0.016	±0.359	±0.212	H10
			±0.079	±0.346	O3
				±0.234	C19

c) Tolerance matrix of ECSA for 34 compounds with low activity

C3	N1	H10	O3	C19	
±0.039	±0.025	±0.043	±0.015	±1.102	C3
	±0.048	±0.007	±0.234	±0.951	N1
		±0.016	±0.419	±0.947	H10
			±0.079	±1.221	O3
				±0.515	C19

d) Tolerance matrix of ECSA for 794 conformations of 72 compounds

C3	N1	H10	O3	C19	
±0.045	±0.027	±0.047	±0.023	±1.159	C3
	±0.048	±0.008	±0.256	±1.131	N1
		±0.017	±0.502	±1.174	H10
			±0.107	±1.440	O3
				±1.488	C19

Thus, in Eq. 4, only the choice of the $a_n^{(j)}$ parameters remained uncertain. Elucidating these required the use of a variable selection method. We developed the GA procedure in the Matlab environment. Each prepared parameter was assigned to Eq. 2 and then solved by the lsqnonlin function within the statistics toolbox in MATLAB in order to obtain the κ_j values of the corresponding model parameters by numerically solving the system of differential equations. The predicted residual error sum of squares (PRESS) for each model was used to determine the fitness of the individual. The PRESS procedure is equivalent to “leave-one-out” cross-validation. It is defined as

$$PRESS_p = \sum_{n=1}^N |A_n^{\text{exp}} - A_n^{\text{calc}}|^2 \quad (5)$$

where A_n^{exp} and A_n^{calc} are the experimental and predicted activities of the n -th molecule of the training set in the leave-one-out cross-validation (LOO-CV) model, respectively. N is the total number of training compounds in the entire data set. P is the number of selected variables used in the regression analysis. The quality of the models was determined by the LOO-CV procedure. q^2 estimates the proportion of variation in the activity that is explained by the predictor. Model validation was done to test the internal stabilities and predictive abilities of the QSAR models. Previously developed QSAR models were evaluated both internally and externally by LOO-CV [54]. Internal validation was necessary to verify the predictivity of the model for the training chemicals and to exclude overfitted or by-chance models. For internal validation, the internal stability of a model was calculated by the following q^2 equation:

$$q^2 = 1 - \frac{\sum_{n=1}^N |A_n^{\text{exp}} - A_n^{\text{calc}}|^2}{\sum_{n=1}^N |A_n^{\text{exp}} - \bar{A}_n^{\text{exp}}|^2} \equiv 1 - \frac{PRESS}{SSY} \quad (6)$$

where \bar{A}_n^{exp} is the average activity of all molecules in the training set. SSY is the sum of the squares of the deviations of the experimental values (A_n^{exp}) from their mean (\bar{A}_n^{exp}). In principle, external validation is the only way to determine the true predictive power of a QSAR model. It is generally considered the most rigorous validation procedure because the compounds in the external test set do not affect model development. The predictive abilities of the models were also confirmed by external validation. For external validation, the activity of each molecule in the test set was predicted using the model developed by the training set [55]. Two different expressions ($q_{\text{ext}1}^2$ and $q_{\text{ext}2}^2$) for the calculation of the external q^2 based on predictions for

external test objects are available, and are calculated as follows [56]:

$$q_{\text{ext}1}^2 = 1 - \frac{\sum_{n=1}^N |A_{n_{\text{test}}}^{\text{exp}} - A_{n_{\text{test}}}^{\text{calc}}|^2}{\sum_{n=1}^N |A_{n_{\text{test}}}^{\text{exp}} - \bar{A}_{n_{\text{training}}}^{\text{exp}}|^2} \quad (7)$$

$$q_{\text{ext}2}^2 = 1 - \frac{\sum_{n=1}^N |A_{n_{\text{test}}}^{\text{exp}} - A_{n_{\text{test}}}^{\text{calc}}|^2}{\sum_{n=1}^N |A_{n_{\text{test}}}^{\text{exp}} - \bar{A}_{n_{\text{test}}}^{\text{exp}}|^2} \quad (8)$$

where N is the number of tested molecules, and A_n^{exp} and A_n^{calc} are the actual and predicted activities of the n -th molecule in the test set, respectively. $\bar{A}_{n_{\text{training}}}^{\text{calc}}$ and $\bar{A}_{n_{\text{test}}}^{\text{calc}}$ are the average experimental activities of all molecules in the training and test sets, respectively. Both summations are performed over all molecules in the test set. Thus, the q^2 value is indicative of the predictive power of the current model for the external test set. They differ from each other only in regard to the mean used in the denominator to calculate the sum of squares.

Once we had calculated and generated the descriptor pool by EMRE software, we reduced the original pool to a more manageable size and then selected a number of optimal descriptor subsets to build a set of models using GA. The predictive abilities of the models were tested with a set of molecules (the prediction set) that were not used during the model-building process.

EC-GA is a sophisticated hybrid approach that combines the EC method for pharmacophore identification and bioactivity prediction with GA for variable selection to achieve a powerful optimization method. Since all energetically reasonable conformations of all compounds are used in both pharmacophore identification and biological activity prediction, this study represents the application of 4D-QSAR to 1,4-DHP derivatives. From this point of view, a particular strength of the 4D EC-GA method is that it provides an expression for the conformational ensemble sampling of compounds while weighting them by Boltzmann contribution as a fourth dimension.

Results and discussion

The predictive abilities of the resulting models were evaluated based on the squared correlation coefficients of the test compound. Although all the models obtained had high R^2 values, this does not imply that all the models had high predictive powers. We cannot gauge the predictivity of a model for new chemicals (they are “unknown” during

model development) without using cross-validation; thus, such validation gives no assurance of the predictive power of the model.

The goal of this study was to construct a 4D-QSAR model for 72 1,4-DHP compounds using the EC-GA method. This method allows for both pharmacophore identification and bioactivity prediction. The chemical structures and experimental activities obtained from the literature (pIC_{50}) [46, 47], the activities calculated by this method, and the numbers of conformers for the compounds are listed in Table 1. The underlined compounds in Table 1 correspond to the 22 compounds of the test set, which were randomly selected from the original 72 compounds.

After performing semiempirical calculations for and conformational analyses of the compounds, 794 ECMCs for 72 compounds were created by the EMRE program, as shown in Fig. 1 for compound 72.

To find the pharmacophore, 72 1,4-DHP molecules were arbitrarily assigned to categories of “high activity” or “low activity.” The compounds with $pIC_{50} \geq 8.57$ were classified as high-activity compounds (38 compounds), and molecules with $pIC_{50} < 8.57$ were considered to be low-activity compounds (34 compounds). The lowest-energy conformation of compound 72, which is the most active one in the series, was chosen as a template to compare with the other ECMCs of the conformers. This allows us to obtain the tolerances and a new common ECSA that includes all of the active compounds.

By gradually changing the limits of tolerance for diagonal elements (charges) and nondiagonal elements (bond orders or interatomic distances), we found that the ECSA submatrix that is common to all of the active molecules contains five atoms corresponding to N1, H1, C3, O3 and C19, and this gives the best separation of the active compounds from similar inactive or low-activity ones. The resulting ECSA and its tolerance values for the compounds with both high and low activities are given in Table 2. The corresponding five atoms of the pharmacophore obtained in this way are shown in white. The first submatrix in Table 2 corresponds to the ECSA (pharmacophore atoms) for the template compound. The second and third matrices show the tolerance values of the lowest energy conformers for the 38 compounds with high activity and 34 compounds with low activity, respectively. Then we calculated the tolerance values for all of the conformers (794) for all of the compounds (72) without any constraints on tolerance limits, and obtained the fourth tolerance matrix. The tolerance matrix for less active compounds does not fit the tolerance matrix for more active compounds. As seen in Table 2, the tolerance values of compounds with high activities are usually lower than those of compounds with low activities. For example, the tolerance values for distance between the H10 and

O3 atoms for high-activity and low-activity compounds are ± 0.359 and ± 0.419 , respectively.

The values of n_1 , n_2 , n_3 and n_4 were found to be 37, 1, 3, and 31, respectively. In this way, the value of parameters expressing the probability of feature realization are adequate: $P_\alpha = 0.905$, $\alpha_a = 0.890$. The analysis of the pharmacophore shows that the presence of the N1, H1, C3, O3 and C19 atoms is a key component of activity within the specifics of the biochemical interaction mechanism for 1,4-dihydropyridine derivatives. The N1, H10 and O3 Pha atoms are considered to be among the hydrophilic regions that are responsible for hydrogen-bond interactions between ligand and receptor.

For the compounds that have the pharmacophore, we can predict calcium channel antagonist activity values using Eq. 4. The data set of 72 molecules was randomly divided into two groups; a training set used for model generation and a prediction set used to evaluate the generated model; these sets consisted of 50 and 22 molecules, respectively. Test compounds were not included in the model generation.

Due to the diversity of the molecules studied in this work, over 150 descriptors were calculated for each conformer of the 72 compounds. Each descriptor encoded different aspects of the molecular structure and consisted of electronic, thermodynamic geometrical and quantum-chemical descriptors.

The operators of the GA tools menu were chosen as follows: the larger the initial population created, the more likely the best solution from it will be closer to optimal, but at a cost of increased execution time. Therefore, a value of 1000 was used as the population size. A stochastic uniform selection function was used. The number of both generations and iterations was set to 500. The fitness of each chromosome, which is the predicted residual error sum of squares (PRESS) of the model, is expressed by Eq. 5. The one-point crossover method was applied to the parent's generation with a crossover fraction of 85%. The mutation rate was fixed at 1.5%; i.e., mutations should be applied to 1.5% of the total population size. The Matlab codes we devised include not only the various general steps of GA—such as the selection of optimal individuals in a population, crossover, and mutation—but also a combination with the LOO-CV method to calculate antagonist activity. The fundamental process of the GA program is illustrated in module 3 of Fig. 1. In order to show the performance of this study for variable selection, the EC-GA method was applied to the inhibitory activity of 1,4-DHPs (known calcium channel antagonists).

To check the robustness of the EC-GA method and obtain a truly predictive and validated model, the original dataset was randomly divided into several training and test sets, with highly predictive models providing q^2 values of greater than 0.7 for the training sets and R^2 values of greater

than 0.6 for the test sets. Multiple QSAR models were generated independently for all training sets and validated using the test sets. Here, the training and test sets contained 72 and 0, 50 and 22, 55 and 17, 45 and 27 compounds, respectively.

Table 3 presents the eight best models selected using the R^2 , q^2 (internal), q_{ext1}^2 and q_{ext2}^2 values from multiple EC-GA models as the criteria. The number of descriptors for the nine best models varied between 9 and 14. In model 1, the first training set contains all the molecules, so there is no test set to validate the model. Figure 3 shows experimental vs. calculated activity values for the training and test sets using these models.

Relying on internal validation methods, such as leave-one-out (LOO) cross-validation, was shown to be an inadequate approach [54]. Best-fit models are not the best for external prediction because internal validation tries to fit the compounds in the training set as well as possible and does not take new compounds into account. Thus, it is not correct to assume that internally cross-validated models will automatically be externally predictive. External test set validation is the best way to assess the predictivity of a model. In Table 3, model 5 is shown to have negative q_{ext1}^2 and q_{ext2}^2 values, indicating that the model generated was not meaningful and produced a nonpredictive model. The predictions from this model are worse than those of the other models. As shown in Table 3, all models based on different combinations of compounds and descriptors (i.e., models 1–8, except for model 5) are highly predictive, with the q^2 values being in the range 0.669–0.836. However, the models differed in their ability to make predictions for new compounds, as assessed by the q^2 and q_{exts}^2 parameters. Model 5 thus showed the lowest q_{ext}^2 values, while model 2 showed the highest. The best model (model 2) with the highest predictive power was obtained for the test set with

22 compounds using the optimal number of 10 descriptors (Table 3). Model 9 was constructed using the same data set and parameters as model 2, but taking into account only the lowest-energy conformer (one conformer) of each compound. Model 9, which uses the same conditions as model 2 except for conformation number, has lower regression coefficients than model 2 for both the training set (0.795) and the test set (0.720). Model 9 also has lower internal and external cross-validated q^2 values than model 2. The different statistical parameter values obtained for both of these models show that representing compounds as an ensemble of conformers in a fourth dimension leads to more accurate predictions for novel compounds, as seen in Table 3.

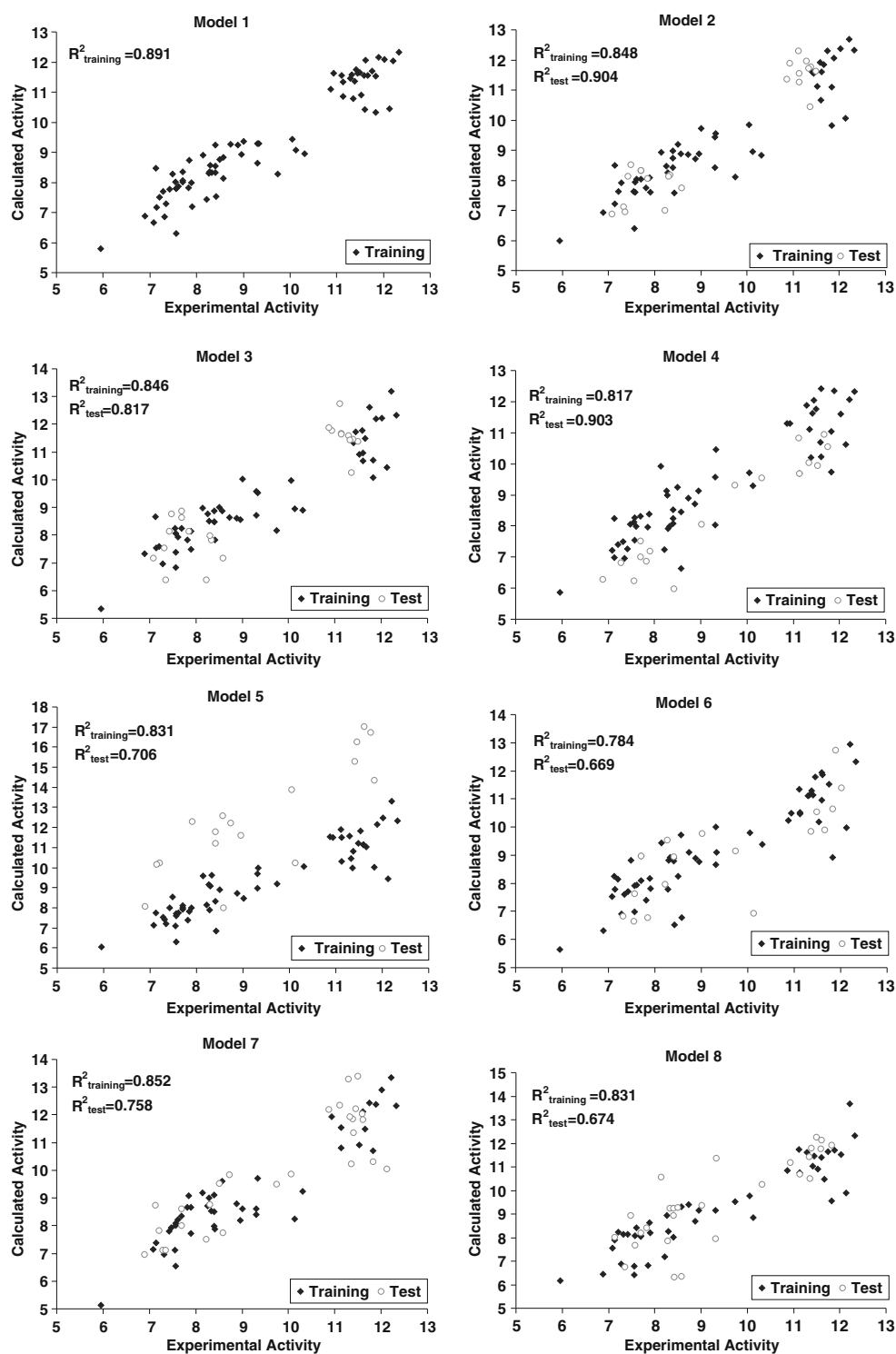
Since the optimum number of variables is not known in advance, several runs are needed to examine the relationship between the predictive power of a model (q^2) and the number of descriptors selected. A GA program was designed to search for the optimal subset among a given number of variables (between 1 and 15). To obtain the optimum number of descriptors, R^2 and q^2 values were plotted against the number of descriptors (Fig. 4). Using the plot, the saturation point was checked, and based on this, ten descriptors was selected as the optimal number and this subset was analyzed further. As a rule of thumb, the data set should be approximately five times larger than the number of descriptors selected in order to achieve good results [57]. As seen in Fig. 4, the best subset of the descriptors is a combination of the ten descriptors that contribute most to activity and describe the influence of ECSA parameter flexibilities. As shown in Fig. 4, the model reaches a steady state after ten variables, and any new variable added after that is redundant. For a training set consisting of 50 compounds, we should have no more than ten descriptors in each model to avoid overfitting the data.

Table 3 Descriptive statistics of the 1,4-DHP data sets used in this study^a

Model no.	N_{TR}	N_{TS}	NOD	$R_{training}^2$	SE	R_{test}^2	SE	q^2	q_{ext1}^2	q_{ext2}^2
1	72	0	14	0.891	0.039	-	-	0.836	-	-
2	50	22	10	0.848	0.056	0.904	0.069	0.780	0.860	0.860
3	50	22	10	0.846	0.057	0.817	0.096	0.767	0.738	0.738
4	55	17	11	0.817	0.059	0.903	0.080	0.707	0.610	0.610
5	55	17	11	0.831	0.056	0.706	0.140	0.764	-3.340	-3.341
6	55	17	11	0.784	0.064	0.614	0.161	0.669	0.523	0.513
7	45	27	9	0.852	0.059	0.758	0.098	0.712	0.712	0.682
8	45	27	9	0.831	0.063	0.674	0.114	0.713	0.560	0.560
9	50	22	10	0.795	0.065	0.720	0.118	0.673	0.676	0.676

^a N_{TR} and N_{TS} : number of data values in the training set and test set, respectively; NOD: number of descriptors used in the models; $R_{training}^2$ and R_{test}^2 : regression coefficients of the training set and test set, respectively; SE: standard error; q^2 : cross-validated regression coefficient; q_{ext1}^2 , q_{ext2}^2 : external validations in the leave-one-out cross-validation

Fig. 3 Calculated vs. experimental activities for eight EC-GA models. Points for the training and test sets are denoted by filled diamonds and blank circles, respectively



A plot of R^2 and q^2 against the number of descriptors (Fig. 4)—which provides guidance regarding the optimal number of descriptors to retain—suggests that the model with ten descriptors performed best, as denoted by the dashed line.

The descriptors selected with GA are listed in Table 4. A careful examination of the descriptors in Table 4 suggests

that electronic, geometric, thermodynamic and solvation parameters are the most influential parameters in the antagonist activities of these compounds.

In Table 4, $a^{(1)}$ is the nucleophilicity index ($1/W$). Parr et al. [58] defined the electrophilicity index (W). On the basis of the assumption that electrophilicity and nucleophilicity are inversely related to each other, Chattaraj et al. [59]

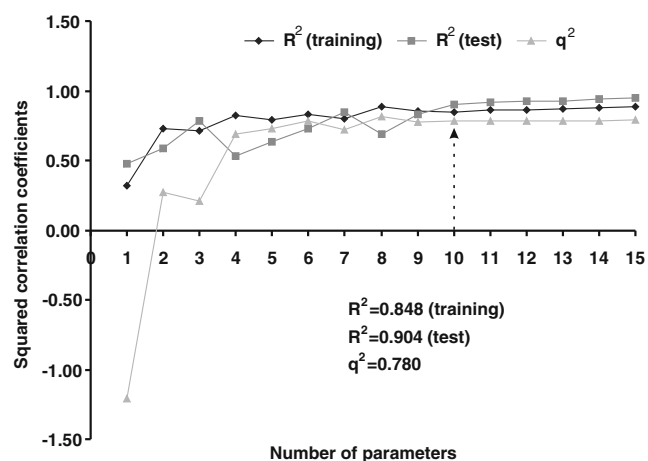


Fig. 4 Plots of R^2 (training), R^2 (test) and cross-validated q^2 against the optimal number of descriptors

suggested a multiplicative inverse of the electrophilicity index ($1/W$), as well as an additive inverse [60]. $a^{(2)}$ and $a^{(3)}$ correspond to two different electronic parameters, $Q(\text{plus})$ and dipole x , respectively. $Q(\text{plus})$ is the magnitude of the most positively charged hydrogen atom in the molecule [61]; it is a hydrogen bonding acidity component that characterizes solvent effects [62]. Dipole x is the x component of the dipole moment in the x -direction, and it is an important molecular descriptor that depicts the directional electronic energy of the compound, which is very useful for postulating ligand interactions [63]. When dipole–dipole interactions are involved in ligand–receptor interactions, the dipole moment on the x -axis plays an important role in the antagonist activity of these compounds. $a^{(4)}$ is the vibration energy of a molecule at absolute zero (0 K). ZPE is the quantum mechanical effect that is a consequence of the uncertainty principle. $a^{(5)}$ is the fundamental frequency of vibration. The lowest vibration frequency of the lowest-energy conformer for the reference compound is only 1.48 cm^{-1} , so it can easily be distorted in

the interaction between the ligand and the receptor. The $a^{(6)}$, $a^{(7)}$, $a^{(8)}$ and $a^{(9)}$ parameters are geometrical parameters that indicate the effect of molecular flexibility on the activity of the reference compound, as shown in Fig. 5. $a^{(6)}$ is the orthogonal distance from the C7 atom to the O3–C3–H10 plane (in green) composed of Pha atoms. $a^{(7)}$ (in violet red) is the angle (in radians) between the C17–O3–H10 plane and the C2–C16 line. The $a^{(8)}$ and $a^{(9)}$ parameters shown in blue indicate interatomic distances in Fig. 5. The C16 atom refers to the farthest atom from the O5 atom in substituent R^1 , except for hydrogen atoms bonding to carbon atoms as the farthest atoms. Similar to C16 atom, the C15 atom is the farthest atom from the O6 atom in the substituent R^2 . For the sake of simplicity, hydrogen atoms attached to carbon atoms are not shown in Fig. 5. Pha atoms are shown in yellow circles. $a^{(10)}$ is the solvation energy E_{aq} (kcal mol^{-1}). Since any changes in the solvation energy will affect the energy required to remove the molecule from the solution upon entering the binding site, $a^{(10)}$ may be considered a predictor of activity. Solvation energy reflects the compound's ability to form noncovalent interactions with the binding site, to dissolve and persist in water, or to permeate the phase interfaces.

Activity depends exponentially on S ($A \sim e^{-S}$) [51]. If the product of the parameter and the kappa values is positive then it shows an APS group; otherwise (if the product is negative) it shows AG. $a^{(1)}$, $a^{(2)}$, $a^{(4)}$, $a^{(5)}$, $a^{(6)}$, $a^{(7)}$ and $a^{(8)}$ are APS groups. The $a^{(9)}$ parameter is AG. $a^{(3)}$ and $a^{(10)}$ have positive or negative values in different conformers of the same compound: they are not only AGs, but also APS.

Model 2, based on a training set of 50 molecules, was validated by performing predictions on 22 test set compounds. Using the κ_j coefficients obtained from the training set by the lsqnonlin function, the activities of the test compounds were calculated using Eq. 3. The goodness-of-fit of a model is a measure of how well the model accounts for the variance in the response in the training set, and is given by the correlation coefficient and the standard error

Table 4 Optimum ten molecular parameters selected with GA and κ_j values used to calculate the activities of 1,4-DHP derivatives

$a_{mi}^{(j)}$	Molecular parameters	K_j value
$a^{(1)}$	Nucleophilicity index (eV^{-1})	11.065
$a^{(2)}$	$Q(\text{plus})$ (e^+)	3.690
$a^{(3)}$	Dipole x (Debye)	−0.006
$a^{(4)}$	Zero-point vibrational energy (kcal/mol)	0.003
$a^{(5)}$	Fundamental frequency of vibration (cm^{-1})	0.013
$a^{(6)}$	Orthogonal distance from the C7 atom to the O3–C3–H10 plane (Å)	0.018
$a^{(7)}$	The angle between the C17–O3–H10 plane and the C2–C16 line (radians)	0.149
$a^{(8)}$	C6–C16 distance (Å)	0.034
$a^{(9)}$	C6–C15 distance (Å)	−0.062
$a^{(10)}$	E_{aq} (kcal/mol)	0.004

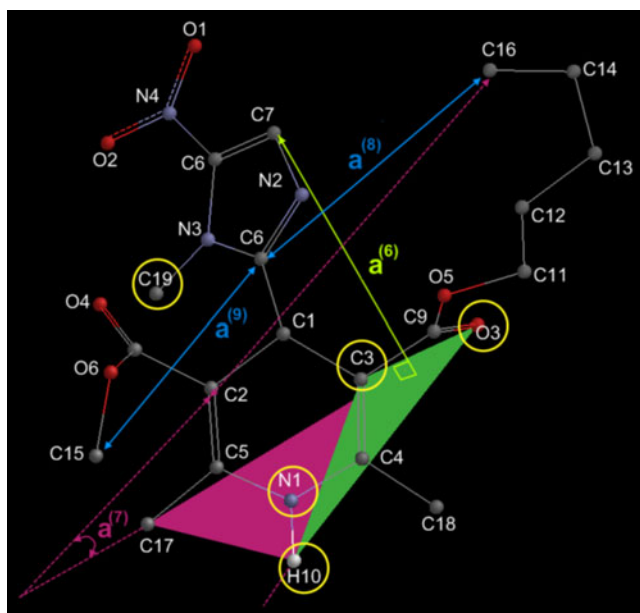


Fig. 5 Representation of $a^{(6)}$, $a^{(7)}$, $a^{(8)}$ and $a^{(9)}$ parameters. In the figure, $a^{(6)}$ is the orthogonal distance from the C7 atom to the O3–C3–H10 plane (in green). $a^{(7)}$ is the angle (in radians) between the C17–O3–H10 plane (in violet red) and the C2–C16 line. The C16 atom refers to the farthest atom from the O5 atom in substituent R¹, except for hydrogen atoms that are bonded to these farthest carbon atoms. $a^{(8)}$ and $a^{(9)}$ parameters (in blue) indicate interatomic distances for C6–C16 and C6–C15, respectively. Similar to the C16 atom, the C15 atom is the farthest atom from the O6 atom in the substituent R². Pha atoms are shown in yellow circles

of the estimate. The model for the training set obtained by the optimum ten parameters gave a satisfactory squared correlation coefficient of 0.848 with a standard error of 0.056 and a LOO-cross-validated q^2 value of 0.780, indicating good predictive ability. Plots of predicted activities versus the experimental values of antagonist activities are shown individually for the training and test sets in Fig. 3. Our results also show that the model with the highest predictive power for the external validation of the test set possesses high values of both q_{ext1}^2 (0.860) and q_{ext2}^2

(0.860). The squared correlation coefficient value (R^2) is 0.904, with a standard error of 0.069.

To estimate the individual influence of each of the ten molecular descriptors on antagonist activity, the E statistics technique [64, 65], which is an important practical tool, was easily implemented for the developed EC-GA model. Each descriptor was ignored once, and its influence was evaluated with the remaining nine descriptors. To see which descriptor contributed the most in a given component, we considered the E , q^2 , $R_{training}^2$, R_{test}^2 , q_{ext1}^2 and q_{ext2}^2 values that are displayed in Table 5. This procedure was then repeated for all of the descriptors present in the model. The following formula was employed to calculate the E value (Eq. 10) [39]:

$$E = \frac{PRESS_P}{PRESS_{P-1}} \quad (10)$$

$$PRESS_{P-1} = \sum_{n=1}^{P-1} |A_n^{exp} - A_n^{calc}|^2 \quad (11)$$

where A_n^{calc} represents the predicted and A_n^{exp} the experimental activities in the LOO cross-validation model, with P parameters being used (in this case $P=10$).

The more a descriptor is related to the model, the greater the overall performance drop when it is omitted. Thus, Table 4 indicates that the angle (in radians) between the C17–O3–H10 plane and the C2–C16 line, $a^{(7)}$, contributes more significantly than all other descriptors for this model. Although the E value is greatest for $a^{(7)}$ (1.129), the omission of this descriptor results in the greatest decrease in the cross-validated q^2 value to 0.113, showing that it is very important to the model. Also, when it is omitted, relatively low R_{test}^2 , q_{ext1}^2 and q_{ext2}^2 values indicate the decline in overall model performance. Solvation energy, $a^{(10)}$, is the second most influential descriptor, and it has the lowest E value. Omission of the solvation energy also leads to a considerable decrease in the cross-validated q^2 value to

Table 5 E , q^2 , $R_{training}^2$, R_{test}^2 , q_{ext1}^2 and q_{ext2}^2 values showing the contribution of each descriptor to the performance of the model of 1,4-DHP antagonist activity

	$a^{(1)}$	$a^{(2)}$	$a^{(3)}$	$a^{(4)}$	$a^{(5)}$	$a^{(6)}$	$a^{(7)}$	$a^{(8)}$	$a^{(9)}$	$a^{(10)}$
E	0.897	0.917	0.550	0.712	1.044	1.012	1.129	1.019	0.734	0.401
$R_{training}^2$	0.835	0.834	0.748	0.822	0.860	0.852	0.873	0.852	0.807	0.643
R_{test}^2	0.712	0.797	0.686	0.726	0.794	0.834	0.743	0.764	0.754	0.469
q^2	0.747	0.752	0.587	0.681	0.782	0.775	0.113	0.777	0.690	0.433
q_{ext1}^2 [a]	0.668	0.793	0.681	0.670	0.758	0.823	0.711	0.683	0.748	0.418
q_{ext2}^2 [a]	0.666	0.792	0.679	0.669	0.757	0.822	0.709	0.682	0.747	0.415

^a q_{ext1}^2 and q_{ext2}^2 are external validations in the leave-one-out cross-validation for ten parameters, as calculated by Eqs. 7 and 8

0.433, as well as in $R_{training}^2$, R_{test}^2 , q_{ext1}^2 and q_{ext2}^2 . Since $a^{(7)}$ and $a^{(10)}$ play major roles in the model's predictive ability, ignoring these descriptors leads to a greater loss of predictive ability than ignoring other descriptors in the model. Individually, the inclusion of $a^{(6)}$, $a^{(2)}$ and $a^{(5)}$ in the model has only a minor effect on the model. While $a^{(5)}$ leads to negligible increases in prediction power (q^2), its omission results in a considerable drop in the R_{test}^2 , q_{ext1}^2 and q_{ext2}^2 values. Consequently, the optimum ten parameters revealed by the developed EC-GA model clearly indicate the importance of electronic, geometric and thermodynamic descriptors in the biological activity of 1,4-dihydropyridines.

QSAR techniques play an important role in the rational design of drugs. Two commonly used techniques are CoMFA and CoMSIA analyses [66, 67]. They have been successfully applied to the design of several new molecules [68]. However, conformation determination and molecular alignment are factors that affect the successful development of a CoMFA or CoMSIA model. Pharmacophore models are used when some active compounds have been identified but the three-dimensional structure of the receptor is unknown [21]. The active compounds are superimposed to find the best overlay of the corresponding features and hence to provide a pharmacophore model. The main difficulty in pharmacophore generation is handling conformational flexibility, since the active conformations of the molecules are usually unknown. Most 3D QSAR methods contain just one conformation for each compound. Therefore, 3D QSAR methods probably will not offer a significant advantage unless the conformational flexibility of the molecule is taken into account. The EC-GA method outlined in this study is based on a rather full description of the conformers of the molecular structures that interact with the biological target. At the same time, this approach has an obvious advantage: it differentiates between molecular configurations and takes into account their conformational features, because conformations are represented by different three-dimensional matrices. Unlike some other methods, it allows for the quantitative prediction of activity, provided the experimental data are available to form a sufficiently representative model. In this way, this method, by complementing the pharmacophore with parameterized out-of-pharmacophore groups and other structural parameters, raises the concept of the pharmacophore from a qualitative to a quantitative tool for bioactivity prediction. The parameters of the pharmacophore, together with the adjustable constants (κ_j) of the parameters in Eq. 3, can be used for the nonexperimental screening of new compounds. The numerical activity values of compounds may also be improved by including a large number of conformations for each compound.

Model interpretation is a very important component of QSAR studies; without it, there is only a restricted capacity

to modify compounds. Modeling algorithms are widely used as a tool for understanding and interpreting. Nonlinear models are generally thought to be difficult to interpret, and these modeling algorithms or models are sometimes referred to as blackbox models [69, 70]. To find new potentially active compounds using the EC-GA method, structural information about the new compound is needed. In general, in the EC-GA method, after performing quantum-chemical calculations to obtain electronic, geometric and thermodynamic features of the structure of the novel molecule, its bioactivity is predicted using the selected parameters and corresponding κ coefficients in Eq. 4 for the related molecule, which then leads to a consideration of possible structural modifications to the molecule. However, some preliminary estimates can be made by analyzing the values and the signs of the κ coefficients. This is the only approach that allows us to investigate how the structure could be modified to enhance activity, since bioactivity exponentially depends on the selected parameters. Otherwise, if we made modifications based on structural fragments, it would not be possible to determine which substituents would lead to higher activity.

Using the theoretical formula (Eq. 4) of the EC method together with the genetic algorithm for predicting calcium channel antagonist activity, we are able to identify new potentially active 1,4-dihydropyridines containing the substituted nitroimidazolyl group. The 4D-QSAR paradigm (EC-GA) combines molecular modeling and QSAR analysis in order to achieve the construction of (in this application) quantitative models of calcium channel antagonists as a function of conformation ensemble.

The first important observation obtained from the pharmacophore identification of 1,4-dihydropyridines is that the substituted nitroimidazolyl ring and the pyridine ring that make up the DHP structure (the phenyl ring and the pyridine ring) are necessary for calcium channel antagonist activity (Fig. 5). Also, the presence of the O3 atom belonging to the carbonyl group attached to the C3 atom of the pyridine ring is a key component of the activity, and can be used to screen new compounds with given tolerance values (Table 2). The pharmacophore consists of two hydrophobic atoms (C3 and C19), two hydrogen bond acceptors (N1 and O3), and one hydrogen bond donor (H10). From this study, one can conclude that the interaction between a molecule showing calcium channel antagonist activity and its receptors takes place through the hydrophobic and donor–acceptor parts of the molecule. Our results show that increasing the fundamental vibration frequency and geometrical parameters such as the orthogonal distance from the C7 atom to the O3–C3–H10 plane, the angle between the C17–O3–H10 plane and the line C2–C16, and the C6–C16 distance decreases activity. Hence, modified compounds with lower values for these param-

ters may show higher activity ($q^2=0.780$). When evaluating the results, it is important to bear in mind that this method not only produces good predictive models but also allows the most influential variables to be discerned. The best model (model 2) was obtained with the genetic algorithm, showing the importance of thermodynamic, electronic and geometric parameters. Therefore, the activity features (pharmacophore) and molecular descriptors represent valuable information for the design of new active compounds or for the modification of pre-existing molecules when used together in the 4D-QSAR models.

Conclusions

We developed a new structure-based 4D-QSAR method called EC-GA that combines the EC and GA methods. In this paper, the EC-GA model developed to predict the calcium channel antagonist activity of 1,4-DHPs allows the molecular properties/features that play an important role in governing the differences between the activities of different molecules to be discerned. The EC-GA method—given its advantageous use of the genetic algorithm and activity equation (Eq. 4)—was shown to be an efficient method to apply to this complex problem, with nonlinear relationships between conformers of molecular structures and biological activities in the QSAR data. One of our aims was to consider the conformational flexibilities of these compounds, in terms of both pharmacophore identification and bioactivity prediction. The probability distribution of the conformers for bioactivity prediction was approximated from the Boltzmann energy distribution. A comprehensive pharmacophore identification, molecular descriptor and activity calculation by a new software package (EMRE) was successfully applied to a series of 1,4-DHPs. The above results show that the predicted biological activity values are in very good agreement with the experimental values. The following conclusions can be drawn from this study:

- (1) The EC-GA method applied to describe the calcium channel antagonist activity of 1,4-DHPs reveals that this activity is controlled by a pharmacophore that consists of five atoms with certain electronic and geometric characteristics: one oxygen atom from the carbonyl group, a nitrogen and a hydrogen atom from the dihydropyridine group that are attached to each other, and two carbon atoms. The interpreted pharmacophore model reflects functional interactions like H-bonds or ionic transfer interactions.
- (2) The proposed models confirmed that the interactions between the receptors and ligands are dependent on the properties of the pharmacophore, not the whole ligand

molecule. This pharmacophore can be used to screen molecular databases for similar action mechanisms.

- (3) The relationship between the structural parameters and the activity factors of the studied calcium channel antagonists can be accurately described as nonlinear. Moreover, the ten selected descriptors represent the features of the compounds most responsible for their binding behavior. The geometric, electronic and thermodynamic parameters and the conformations of the compounds are closely related to the biological activities of these compounds.

The developed EC-GA model also possesses promising predictive ability, as discerned by testing it on an external test set, and could be useful for elucidating the relationship between compound structures as a function of conformation ensemble and biological activity, and for facilitating the design of more potent calcium channel blockers. The developed EC-GA model displays consistent descriptor contributions to activities and statistical significances in the training (50 compounds) and test (22 compounds) sets. The resulting QSAR model was found to be statistically significant with respect to training ($R^2=0.848$), testing ($R^2=0.904$), cross-validation ($q^2=0.780$), and external validation ($q_{ext1}^2=0.860$, $q_{ext2}^2=0.860$) values. The robustness of the model for the training and test sets referred to the stability of the selected parameters and thus the stability of its predictions. The parameters suggested by model 2 in this study could be used to identify and predict the biological effects of other molecules that have not yet been synthesized or tested. The EC-GA method was shown to be efficient for determining the Pha and the activities of calcium channel antagonists. Although the 4D-QSAR approach used in this study could not directly indicate the receptor structure of calcium channel antagonists, results demonstrated that statistically predictive 1,4-DHPs models could be generated. In the future, a more comprehensive study of the protein structures of calcium channel antagonist receptors and the docking of active molecules to them may explain their interaction mechanism.

Acknowledgement This project was financially supported by the Scientific Technical Research Council of Turkey (TUBITAK, Grant No. 107T385) and the Research Fund of Erciyes University (BAP, Project Number: FBD-09-928). The authors are grateful to Mustafa Yıldırım, Fatih Çopur and Serkan Şahin for their valuable suggestions.

References

1. Itai A, Tomioka N, Kato K (1995) In: Fujita T (ed) QSAR and drug design: new developments and applications. Elsevier, Amsterdam
2. Thomas G (2003) Fundamentals of medicinal chemistry. Wiley-Blackwell, Weinheim

3. Şafak C, Şimşek R (2006) Fused 1,4-dihydropyridines as potential calcium modulatory compounds. *Mini Rev Med Chem* 6:747–755
4. Fosshem R (1986) Crystal structure of the dihydropyridine Ca^{2+} antagonist felodipine. Dihydropyridine binding prerequisites assessed from crystallographic data. *J Med Chem* 29:305–307
5. Triggle DJ, Lings DA, Janis RA (1989) Ca^{2+} channel ligands: structure-function relationships of the 1,4-dihydropyridines. *Med Res Rev* 9:123–180
6. Jiang JL, Li AH, Jang SY, Chang L, Melman N, Moro S, Ji X, Lobkovsky EB, Clardy JC, Jacobson KA (1999) Chiral resolution and stereospecificity of 6-phenyl-4-phenylethynyl-1,4-dihydropyridines as selective A3 adenosine receptor antagonists. *J Med Chem* 42:3055–3065
7. Rhee AM, Jiang JL, Melman N, Olah ME, Stiles GL, Jacobson KA (1996) Interaction of 1,4-dihydropyridine and pyridine derivatives with adenosine receptors: selectivity for A3 receptors. *J Med Chem* 39:2980–2989
8. Triggle DJ (2003) 1,4-Dihydropyridines as calcium channel ligands and privileged structures. *Cell Mol Neurobiol* 23:293–303
9. Kahraman P, Turkay M (2007) Classification of 1,4-dihydropyridine calcium channel antagonists using the hyperbox approach. *Ind Eng Chem Res* 46:4921–4929
10. Mohajeri A, Hemmateenejad B, Mehdipour A, Miri R (2008) Modeling calcium channel antagonistic activity of dihydropyridine derivatives using QTMS indices analyzed by GA-PLS and PC-GA-PLS. *J Mol Graph Model* 26:1057–1065
11. Yao X, Liu H, Zhang R, Liu M, Hu Z, Panaye A, Doucet JP, Fan B (2005) QSAR and classification study of 1,4-dihydropyridine calcium channel antagonists based on least squares support vector machines. *Mol Pharm* 2:348–356
12. Takahata Y, Costa MCA, Gaudio AC (2003) Comparison between neural networks (NN) and principal component analysis (PCA): structure activity relationships of 1,4-dihydropyridine calcium channel antagonists (nifedipine analogues). *J Chem Inf Comput Sci* 43:540–544
13. Schleifer KJ, Tot E (2002) CoMFA, CoMSIA and GRID/GOLPE studies on calcium entry blocking 1,4-dihydropyridines. *Quant Struct Act Relat* 21:239–248
14. Safarpour MA, Hemmateenejad B, Miri R, Jamali M (2003) Quantum chemical-QSAR study of some newly synthesized 1,4-dihydropyridine calcium channel blockers. *QSAR Comb Sci* 22:997–1005
15. Niculescu SP (2003) Artificial neural networks and genetic algorithms in QSAR. *J Mol Struct THEOCHEM* 622:71–83
16. Cronin MTD, Schultz TW (2003) Pitfalls in quantitative structure–activity relationships (QSARs) for predicting toxicity. *J Mol Struct THEOCHEM* 622:39–52
17. Schultz TW, Cronin MTD (2003) Essential and desirable characteristics of ecotoxicity QSARs. *Environ Toxicology Chem* 22:599–607
18. Kubinyi H, Folkers G, Martin YC (eds) (1998) 3D QSAR in drug design: recent advances. Kluwer, Dordrecht
19. Hopfinger AJ, Wang S, Tokarski JS, Jin B, Albuquerque M, Madhav PJ, Duraiswami C (1997) Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J Am Chem Soc* 119:10509–10524
20. Becker OM, Levy Y, Ravitz O (2000) Conformation spaces, and bioactivity. *J Phys Chem B* 104:2123–2135
21. Langer T, Hoffmann RD (eds) (2006) Pharmacophores and pharmacophore searches. Wiley-VCH, Weinheim
22. Guner OF (2002) History and evolution of the pharmacophore concept in computer-aided drug design. *Curr Top Med Chem* 2:1321–1332
23. Dimoglu AS, Vlad PF, Shvets NM, Coltsa MN, Guzel Y, Saracoglu M, Saripinar E, Patat S (1995) Electronic-topological investigations of the relationships between chemical structure and ambergris odor. *New J Chem* 19:1217–1226
24. Saripinar E, Guzel Y, Patat S, Yildirim I, Akcamur Y, Dimoglu A (1996) Electron-topological investigation of the structure–anti-tubercular activity relationship of thiosemicarbazone derivatives. *Arzneim Forsch (Drug Res)* 46:824–828
25. Guzel Y, Saripinar E, Yildirim I (1997) Electron-topological (ET) investigation of structure–antagonist activity of a series of dibenzo [a,d]cycloalkenimines. *J Mol Struct THEOCHEM* 418:83–91
26. Rosines E, Bersuker IB, Boggs JE (2001) Pharmacophore identification and bioactivity prediction for group I metabotropic glutamate receptor agonists by the electron-conformational QSAR method. *Quant Struct Act Relat* 20:327–333
27. Makkouk Al H, Bersuker IB, Boggs JE (2004) Quantitative drug activity prediction for inhibitors of human breast carcinoma. *Int J Pharm Med* 18:81–89
28. Marenich AV, Yong PH, Bersuker IB, Boggs JE (2008) Quantitative antidiabetic activity prediction for the class of guanidino- and aminoguanidinopropionic acid analogs based on electron-conformational studies. *J Chem Inf Model* 48:556–568
29. Bersuker IB, Bahceci S, Boggs JE, Pearlman RS (1999) A novel electron-conformational approach to molecular modeling for QSAR by identification of pharmacophore and anti-pharmacophore shielding. *SAR QSAR Environ Res* 10:157–173
30. Bersuker IB (2008) QSAR without arbitrary descriptors: the electron-conformational method. *J Comput Aided Mol Des* 22:423–430
31. Eriksson L, Andersson PL, Johansson E, Tysklind M (2006) Megavariate analysis of environmental QSAR data. Part I. A basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD). *Mol Divers* 10:169–186
32. Dudek AZ, Arodz T, Galvez J (2006) Computational methods in developing quantitative structure–activity relationships (QSAR): a review. *Comb Chem High Throughput Screening* 9:213–228
33. Holland JH (1975) Adaptation in artificial and natural systems. University of Michigan, Ann Arbor
34. Terfloth L, Gasteiger J (2001) Neural networks and genetic algorithms in drug design. *DDT* 6:102–108
35. Verma A, Llorca X, Venkataraman S, Goldberg DE, Campbell RH (2010) Scaling eCGA model building via data-intensive computing. In: WCCI 2010 IEEE World Congr on Computational Intelligence, Barcelona, Spain, 18–23 July 2010
36. Jones G (2010) GAPE: an improved genetic algorithm for pharmacophore elucidation. *J Chem Inf Model* 50:2001–2018
37. Reddy AS, Kumar S, Garg R (2010) Hybrid-genetic algorithm based descriptor optimization and QSAR models for predicting the biological activity of tipranavir analogs for HIV protease inhibition. *J Mol Graph Model* 28:852–862
38. Mercader AG, Duchowicz PR, Fernandez FM, Castro EA (2010) Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines. *J Chem Inf Model* 50:1542–1548
39. Guha R, Jurs PC (2005) Interpreting computational neural network QSAR models: a measure of descriptor importance. *J Chem Inf Model* 45:800–806
40. Saripinar E, Geçen N, Sahin K, Yanmaz E (2010) Pharmacophore identification and bioactivity prediction for triaminotriazine derivatives by electron conformational-genetic algorithm QSAR method. *Eur J Med Chem* 45:4157–4168
41. Sahin K, Saripinar E, Yanmaz E, Geçen N (2011) Quantitative bioactivity prediction and pharmacophore identification for benzotriazine derivatives using the electron conformational-genetic algorithm in QSAR. *SAR and QSAR Environ Res*. doi:10.1080/1062936X.2010.548341
42. Yanmaz E, Saripinar E, Sahin K, Geçen N, Çopur F (2011) 4D-QSAR analysis and pharmacophore modeling: Electron

- conformational-genetic algorithm approach for penicillins. *Bioorg Med Chem*. doi:10.1016/j.bmc.2011.02.035
43. Shafiee A, Miri R, Dehpour AR, Soleymani F (1996) Synthesis and calcium-channel antagonist activity of nifedipine analogues containing nitroimidazolyl substituent in guinea-pig ileal smooth muscle. *Pharmaceut Sci* 2:541–543
 44. Miri R, Howlett SE, Knaus EE (1997) Synthesis and calcium channel modulating effects of isopropyl 1,4-dihydro-2,6-dimethyl-3-nitro-4-(thienyl)-5-pyridinecarboxylates. *Arch Pharm Pharm Med Chem* 330:290–294
 45. Miri R, McEwen CA, Knaus EE (2000) Synthesis and calcium channel modulating effects of modified Hantzsch nitrooxyalkyl 1,4-dihydro-2,6-dimethyl-3-nitro-4-(pyridinyl or 2-trifluoromethylphenyl)-5-pyridinecarboxylates. *Drug Dev Res* 51:225–232
 46. Miri R, Dehpour AR, Azimi M, Shafiee A (2001) Synthesis and smooth muscle calcium channel antagonist effect of alkyl, aminoalkyl-1,4-dihydro-2,6-dimethyl-4-nitroimidazole-3,5-pyridine dicarboxylates. *J School Pharmacy Med Sci Univ Tehran* 9:40–45
 47. Miri R, Niknahad H, Vesal G, Shafiee A (2002) Synthesis and calcium channel antagonist activities of 3-nitrooxyalkyl, 5-alkyl-1,4-dihydro-2,6-dimethyl-4-(1-methyl-5-nitro-2-imidazolyl)-3,5-pyridinedicarboxylates. *Il Farmaco* 57:123–128
 48. Wavefunction, Inc. (2006) SPARTAN, v.06. Wavefunction, Inc., Irvine
 49. Bersuker IB (2003) Pharmacophore identification and quantitative bioactivity prediction using the electron-conformational method. *Curr Pharm Des* 9:1575–1606
 50. Bersuker IB, Dimoglo AS (1991) The electron-topological approach to the QSAR problem. In: Lipkowitz KB, Boyd DB (eds) *Reviews in computational chemistry*, 2nd edn. Wiley, New York, pp 423–460
 51. Bersuker IB, Bahceci S, Boggs JE, Pearlman RS (1999) An electron-conformational method of identification of pharmacophore and anti-pharmacophore shielding: application to rice blast activity. *J Comput Aided Mol Des* 13:419–434
 52. Dimoglo AS, Shvets NM, Tetko IV, Livingstone DJ (2001) Electronic-topological investigation of the structure–acetylcholinesterase inhibitor activity relationship in the series of *n*-benzylpiperidine derivatives. *Quant Struct Act Relat* 20:31–45
 53. Pavlov T, Todorov M, Stoyanova G, Schmieder P, Aladjov H, Serafimova R, Mekenyan O (2007) Conformational coverage by a genetic algorithm: saturation of conformational space. *J Chem Inf Model* 47:851–863
 54. Consonni V, Ballabio D, Todeschini R (2009) Comments on the definition of the Q^2 parameter for QSAR validation. *J Chem Inf Model* 49:1669–1678
 55. Damme SV, Bultinck P (2007) Software news and update a new computer program for QSAR-analysis: ARTE-QSAR. *J Comput Chem* 28:1924–1928
 56. Schuurmann G, Ebert RU, Chen J, Wang B, Kuhne R (2008) External validation and prediction employing the predictive squared correlation coefficients test set activity mean vs training set activity mean. *J Chem Inf Model* 48:2140–2145
 57. Topliss JG, Edwards RP (1979) Chance factors in studies of quantitative structure–activity relationships. *J Med Chem* 22:1238–1244
 58. Parr RG, Szentpaly L, Liu S (1999) Electrophilicity index. *J Am Chem Soc* 121:1922–1924
 59. Chattaraj PK, Maiti B (2001) Reactivity dynamics in atom–field interactions: a quantum fluid density functional study. *J Phys Chem A* 105:169–183
 60. Vleeschouwer FD, Speybroeck VV, Waroquier M, Geerlings P, Proft FD (2007) Electrophilicity and nucleophilicity index for radicals. *Org Lett* 9:2721–2724
 61. Cramer CJ, Famini G, Lowrey AH (1993) Use of calculated quantum chemical properties as surrogates for solvatochromic parameters in structure–activity relationships. *Acc Chem Res* 26:599–605
 62. Oliferenko AA, Oliferenko PV, Huddleston JG, Rogers RD, Palyulin VA, Zefirov NS, Katritzky AR (2004) Theoretical scales of hydrogen bond acidity and basicity for application in QSAR/QSPR studies and drug design. Partitioning of aliphatic compounds. *J Chem Inf Comput Sci* 44:1042–1055
 63. Patel DM, Patel NM (2009) QSAR analysis of aminoquinoline analogues as MCH1 receptor antagonist. *J Sci Res* 1:594–605
 64. Livingstone D (1995) *Data analysis for chemists*. Oxford University Press, New York
 65. Wold S (1978) Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* 20:397–405
 66. Cramer R, Patterson D, Bunce J (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 110:5959–5967
 67. Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* 37:4130–4146
 68. Yang GF, Huang X (2006) Development of quantitative structure–activity relationships and its application in rational drug design. *Curr Pharm Des* 12:4601–4611
 69. Fox T, Kriegl JM (2006) Machine learning techniques for in silico modeling of drug metabolism. *Curr Top Med Chem* 6:1579–1591
 70. Helma C, Kazius J (2006) Artificial intelligence and data mining for toxicity prediction. *Curr Comput Aided Drug Des* 2:123–133

Structural models of CFTR–AMPK and CFTR–PKA interactions: R-domain flexibility is a key factor in CFTR regulation

Marian Siwiak · Aleksander Edelman ·
Piotr Zielenkiewicz

Received: 8 November 2010 / Accepted: 4 March 2011 / Published online: 1 April 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Cystic fibrosis (CF), the most common lethal genetic disease among Caucasians, is caused by mutations in cystic fibrosis transmembrane conductance regulator (CFTR). CFTR's main role is to transport chloride ions across epithelial cell membranes. It also regulates many cell functions. However, the exact role of CFTR in cellular processes is not yet fully understood. It is recognized that a key factor in CFTR-related regulation is its phosphorylation state. The important kinases regulating CFTR are cAMP-dependent protein kinase A (PKA) and 5'-AMP-activated protein kinase (AMPK). PKA and AMPK have opposite effects on CFTR activity despite their highly similar structures and recognition motifs. Utilizing homology modeling, *in silico* mutagenesis and literature mining, we supplement available information regarding the atomic-resolution structures of PKA, AMPK and CFTR, and the complexes CFTR–PKA and CFTR–AMPK. The atomic-

resolution structural predictions reveal an unexpected availability of CFTR Ser813 for phosphorylation by both PKA and AMPK. These results indicate the key role of the structural flexibility of the serine-rich R-domain in CFTR regulation by phosphorylation.

Keywords CFTR · AMPK · PKA · Interaction · Molecular modeling · Docking

Abbreviations

CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
ABC	ATP-binding cassette
TMD	Transmembrane domain
NBD	Nucleotide binding domain
PKA	cAMP-dependent protein kinase A
AMPK	5'-AMP-activated protein kinase
RMSD	Root mean square deviation

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1029-0) contains supplementary material, which is available to authorized users.

M. Siwiak · P. Zielenkiewicz
Department of Bioinformatics,
Institute of Biochemistry and Biophysics,
Polish Academy of Sciences,
ul. Pawińskiego 5a,
02-106 Warsaw, Poland

A. Edelman
INSERM U845, Faculté de Médecine, Université Paris Descartes,
Paris, France

P. Zielenkiewicz (✉)
Faculty of Biology, Warsaw University,
Miecznikowa 1,
02-096 Warszawa, Poland
e-mail: piotr@ibb.waw.pl

Introduction

Cystic fibrosis (CF) is caused by mutations in cystic fibrosis transmembrane conductance regulator (CFTR) [1]. The main function of CFTR is its role as a protein kinase-regulated chloride channel, but it is a multifunctional protein. CFTR, located in the apical membranes of secretory epithelial cells, also anchors a dynamic protein macro complex involved in multiple cell functions [2].

Clinical observations have identified a large spectrum of phenotypes of CF, even for patients bearing the same mutation. These observations cannot be solely explained by mutations in CFTR. Modifier genes and/or epigenetic

factors may be potential severity modulating factors. Studies suggest that dynamic interactions within the CFTR interactome—important players in the regulation of cell and epithelial functions—can serve as severity modulation factors as well [3].

CFTR belongs to the ATP-binding cassette transporter (ABC) superfamily of proteins. It is a 1480-amino acid transmembrane protein with a symmetrical structure: a repeat composed of a transmembrane region (TMD) and a nucleotide binding domain (NBD) separated by an intrinsically disordered [4, 5] hydrophilic regulatory domain (R domain). The R domain is a unique feature among ABC transporters. A molecular model of the CFTR structure model was prepared by Riordan's group [6], but it is not complete. During this study we successfully extended this model.

Several regulatory factors affecting CFTR chloride channel activity have been reported, but CFTR chloride channel activity is regulated mostly by protein kinases [7]. The following kinases have been reported to be involved in this regulation: cAMP-dependent protein kinase A (PKA), protein kinase C, Ca^{2+} /calmodulin-dependent kinase, cGMP-dependent kinase, CK2 protein kinase [8], and 5'-AMP-activated protein kinase (AMPK) [9]. This list is likely still incomplete.

The best described is the process of activating CFTR channels via phosphorylation by PKA in the presence of ATP. There are 15 serines that need to be mutated into alanines before CFTR loses its PKA activation sensitivity; none of these serines alone is necessary or sufficient for full PKA response [10]. PKA phosphorylation sites are found primarily in the R region, but are also found within the NBD1 sequence [11].

Understanding the mechanisms underlying the regulation of CFTR by PKA also requires an understanding of the mechanisms underlying the opposite effect of another kinase, AMPK, which inhibits the PKA-dependent activation of CFTR and leads to the inhibition of Cl^- secretion in vitro and in vivo [12].

AMPK forms a heterotrimer consisting of subunits α , β and γ . Subunits β , γ and the C-terminal domain of subunit α form the regulatory part of the protein. The N-terminal catalytic domain of subunit α possesses the kinase functionality. Catalytic and regulatory parts are connected with a flexible linker. It has been shown that the AMPK regulatory part binds to the NBD domains of CFTR [13–16]. The R-domain sequence carries two sites recognized by AMPK: “inhibitory” serines S737 and S768; however, CFTR phosphorylation by AMPK is decreased by only ~80% when those two serines are replaced by alanines [7].

In order to gain insight into the interaction between CFTR and its regulatory kinases, the aim of the present work was to create atomic-resolution models of the complexes of CFTR with PKA and AMPK.

Our results allowed us to draw conclusions about the AMPK–CFTR and PKA–CFTR binding interfaces and to prove at the structural level a hypothesis regarding the role of the flexibility of the R domain in CFTR phosphorylation.

Materials and methods

Sequence analysis

The SwissProt/Trembl [17] database was used as the source of sequences of human 5'-AMP-activated protein kinase (AMPK) subunits $\alpha 1$ (id: Q13131), $\beta 1$ (id: Q9Y478), $\gamma 2$ (id: Q9UGJ0), human PKA (id: P17612) and human cystic fibrosis transmembrane conductance regulator (CFTR) (id: P13569).

Templates for structural modeling were identified as sequence homologs among the proteins deposited in the Protein Data Bank [18] using the BLAST algorithm [19] and the HHPRED program [20]. Pairwise alignments of sequences of the AMPK subunits with the chosen templates were prepared with the Needleman–Wunsch algorithm [21], and are included in the “[Electronic supplementary material](#).”

Identified AMPK $\alpha 1$ catalytic domain templates fulfilled the following conditions: (i) the structure has bound ATP/ADP and a substrate peptide; (ii) the structure is not altered by interactions; (iii) sequence similarity enables homology modeling; (iv) the template structure should have the least number of gaps in the structure.

Using these conditions, the α -catalytic subunit of mouse cAMP-dependent protein kinase (pdb: 1JBP:E [22]) was identified as the best template candidate.

For AMPK $\beta 1$, the structures of the yeast proteins SPCC1919.03c and SIP2 (pdb ids: 2OOX:B [23] and 2QLV:B [24] respectively) were used as templates. The best templates to model AMPK $\gamma 2$ were found to be rat 5'-AMP-activated protein kinase subunit γ -1, yeast hypothetical protein C1556.08c, and nuclear protein SNF4 (pdb ids: 2V8Q:E, 2OOX:G and 2QLV:C, respectively). In cases where more than one template was chosen to model a given subunit, the structures of the templates were aligned using CE [25] and FATCAT [26].

The sequence of the α -catalytic subunit of mouse cAMP-dependent protein kinase (pdb: 2GU8 [27]) exhibits 97.6% identity to the sequence of human PKA in the covered fragment 15–351.

Modeling tools

Loop modeling Loops were modeled by searching the PRODAT database with SYBYL 8.0.

Minimization Models were subjected to energy minimization using the AMBER7 99 force field [28], Gasteiger–Hückel charges, a dielectric constant of 80, and simplex initial optimization followed by Powell minimization terminated at a gradient change of 0.1 kcal/(mol Å), as implemented in SYBYL 8.0.

In silico mutagenesis Side chains of mutated residues were replaced using SYBYL 8.0.

Visualization Figures representing protein structures were prepared using PyMol [29].

CFTR modeling

To model CFTR fragment 1156–1208, a secondary structure prediction was obtained using a Jpred [30] server. We decided to rely on Jpred because other (transmembrane-oriented) tools predicted that the sixth helix ends at residue 1148, which is not likely. The predicted helical fragment (CFTR 1156–1180) was modeled using the sixth helix of the TMD1 domain as a template, and connected with NBD2 by modeled loop (CFTR 1181–1208).

CFTR fragment 807–817 was remodeled using peptide bound to the α -catalytic subunit of mouse cAMP-dependent protein kinase as template substrate. The template peptide's residues RRASI were replaced in silico by the corresponding CFTR residues RRLSQ; the residues flanking this sequence were removed. Rotamers of sidechains had to have: (i) minimal structural overlap with the existing structure; (ii) minimal full atom RMSD to the old residues. As a next step, the structure of the prepared pentapeptide was superimposed on the original CFTR model 810–814 fragment. The residues 807–817 were removed from the R domain, and the loops CFTR 807–809 and CFTR 815–817 that link the prepared 810–814 pentapeptide with the R domain were modeled.

AMPK modeling

Catalytic domain: SWISS-MODEL was used to model AMPK α 1 1–274. The modeling of fragments 289–293 and 302–307 using 1JBP:E as a template and linking all fragments with loops was carried out in SYBYL.

Regulatory part: SWISS-MODEL [31] was used to model the C-terminal regulatory domain of AMPK α 1, while SYBYL 8.0 was used to model subunits β 1 and γ 2. Two AMPK α 1 models with two templates were prepared: (i) AMPK α 1 358–478, using carbon catabolite derepressing protein kinase (pdb: 2QLV:A); (ii) AMPK α 1 398–550, using rat 5'-AMP-activated protein kinase subunit α 1 (pdb: 2V8Q:A). Overlapping fragments were structurally aligned

(RMSD 3.06 Å) then removed. Models were merged, with the missing fragment 469–471 modeled as a loop.

There are more gaps in the C-terminal part of the alignment of AMPK β 1 fragment 73–270 with the SIP2 protein β subunit (pdb: 2QLV:B) than in the corresponding part of the alignment with protein SPCC1919.03c (pdb: 2OOX:B). In the modeling of AMPK β 1 73–157, 2QLV:B was used as a template; in the modeling of fragment 167–270, 2OOX:B was used. The models of fragments 73–157 and 167–270 were then structurally aligned with structure 2QLV:B to obtain their position in the AMPK trimer (as we used 2QLV:A and C to model other subunits of the trimer). The missing linker AMPK β 1 158–166 was modeled as a loop.

AMPK γ 2 was modeled using nuclear protein SNF4 (pdb: 2QLV:C) as template.

Human PKA modeling

The structure of the mouse PKA catalytic [27] subunit was used as the template. Eight residues differentiating mouse PKA from human PKA (T19, S21, Q26, D31, S52, A111, E163, and T335) were replaced in silico in SYBYL 8.0 with the corresponding human PKA residues S, A, H, E, T, P, Q, and S, respectively. The structure obtained was subjected to energy minimization.

Modeling of the CFTR–AMPK complex

The interface region of the regulatory domain CFTR was defined on the basis of experimentally determined interaction sites described in the literature. The surface and structure of the interface region was visually scanned for spacial and hydrophobic complementarity. The identified area was used to manually construct the complex subjected to minimization. The catalytic domain was docked to the CFTR R domain using the previously docked PKA structure as a template. Complex analysis was performed using the Solvent Accessible Surface calculator of Michel Sanner's Molecular Surface algorithm [32] and the Protein Interactions Calculator [33]

Results

AMPK model extension

A database search revealed that the available AMPK model [34] could be extended. Most of the potential templates are fragmentary and contain different fragments that are homologous to AMPK. The protein of the sequence most similar to the catalytic domain of AMPK α 1—protein kinase Chk2—is crystallized as a homodimer, with inhib-

itory T-loops projecting from the bodies of the subunits involved in the formation of the interface [35]. This dimerization-enabling feature, specific to Chk2, disqualified its structure. The 1JBP:E structure we used has a RMSD from the Chk2 structure (with T-loop excluded) of below 2 Å.

When modeling the regulatory part, the 2V8Q structure—which was the most sequentially similar—seemed to be the best template for trimer modeling. However, the 2V8Q: E chain was truncated prior to crystallization, resulting in the deformation of the 190–198 fragment. The 2V8Q structure was excluded from further modeling of the regulatory part of AMPK. Structural alignments demonstrated that the structures 2QLV and 2OOX are almost identical to 2V8Q (their similarities in terms of the RMSDs of the protein backbone atom positions are presented in Table 1 in the “Electronic supplementary material”).

For structure-based interaction predictions, it is of the utmost importance to use continuous structural models of the proteins involved, as any discontinuity may contain potential obstacles or introduce clues for complex formation. All of our models were extended to meet the continuity criterion in the interface areas. The range of extension of the AMPK model is presented in Fig. 1.

CFTR–PKA catalytic domain interaction

Ten residues of CFTR are known to be phosphorylated by protein kinase A (PKA). However, even mutations of all ten of these residues do not prevent the activation of CFTR by PKA [36]. In our model, S813—which is known to have the strongest activating effect among all serines phosphorylated in CFTR—is the most exposed. We were aware that

the R domain is highly unstable, and that the conformation under study was just one of many possible. We confirmed that this conformation allowed the binding of PKA to S813. After docking human PKA to CFTR, we observed that the following H-bonds, responsible for substrate peptide recognition, formed between the recognized pentapeptide of CFTR (810–814) and both PKA and PKA-bound ADP: CFTR Arg810–ADP; CFTR Arg810–PKA Tyr331; CFTR Arg811–PKA Lys169, Glu171; CFTR Ser813–PKA Lys169; CFTR Glu814–PKA Leu199. These bonds are presented in Fig. 2a. Lys169 is a residue in the active site associated with the kinase activity of PKA. The surface position of S813, on flexible linkers, agrees with data showing an immediate CFTR response to PKA activation. S813 is accessible, and its environment allows the rapid binding of PKA. The overall structure of the CFTR–PKA complex is presented in Fig. 3a.

CFTR–AMPK catalytic domain complex formation

Riordan’s de novo modeling [4] of the R domain provides many clusters with different conformations. The conformation presented in the template model used is considered to be one of many that occur naturally in a cell. The R domain is connected to other domains via linkers: 658–679 and 843–847. NMR experiments have shown that the flexibilities of some R-domain fragments decrease after NBD1 binding [5]. In the conformation used, these parts are internally stabilized, so they are not predicted to interact directly with NBD1. There is experimental evidence showing large movements of the R domain vs. the NBDs and TMDs [37]. As the relative position of the R domain with respect to channel-forming domains has not been specified, we treated the linkers as

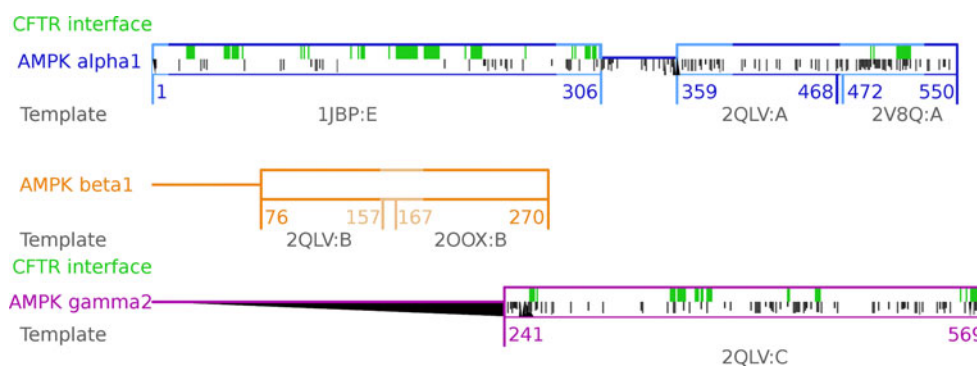
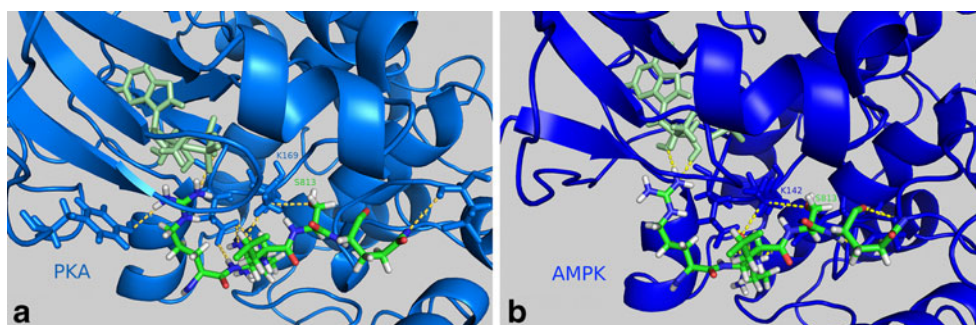


Fig. 1 In-scale visualization of the AMPK modeling effort. *Horizontal boxes* depict modeled fragments. *Horizontal lines* depict fragments of unknown homologs with known structures, probably unstructured fragments. *Lighter box lines* depict fragments of the AMPK model created in this work; *darker lines* depict fragments that were previously modeled. *Green fill* depicts the predicted interface of AMPK with CFTR. *Grayscale shading* depicts differences in sequence between isoforms that interact and do not interact with CFTR

(position-specific score of Needleman–Wunsch alignment of AMPK α 1 vs. AMPK α 2 and AMPK γ 2 vs. γ 1). *Black triangles* depict gaps: those pointing up depict gaps in the presented isoform; those pointing down depict gaps in isoforms that do not interact with CFTR. *Shorter and darker lines* depict different residues. *Numbers below the lines* describe the range of the model: the pdb id of the template structure used to model these fragments is shown in *gray*

Fig. 2 Interactions of CFTR pentapeptide 810–814 with PKA (*left*) and AMPK (*right*). Kinase structures are simplified, and only residues that form H-bonds with the pentapeptide are drawn in detail (yellow dashed lines). CFTR S813 and the lysines at the active site of the kinases are labeled



flexible strings, and the R-domain position was not fixed during the docking of the catalytic domains of AMPK α 1 and PKA. The R domain extracted from the whole CFTR model was scanned for the AMPK recognition motif. Two serines—S737 and S768—were identified. Experimental evidence of the inhibitory role of AMPK after the phosphorylation of S737 and S768 was provided in previous research [37–39]. However, experiments [7] have shown an 80% decrease in AMPK phosphorylation for the CFTR double mutant S737A-S768A. We decided to use relaxed AMPK recognized motif, RRXS/TY, to identify other potential target serines. S813 was identified. We decided to use PKA previously bound to S813 as a template for AMPK α 1 catalytic domain docking. After preparing the R domain–AMPK catalytic domain complex, the R domain was reattached to CFTR.

The R domain with the docked AMPK α 1 catalytic subunit was minimized. The bonds required for substrate peptide binding by protein kinase [22] were satisfied.

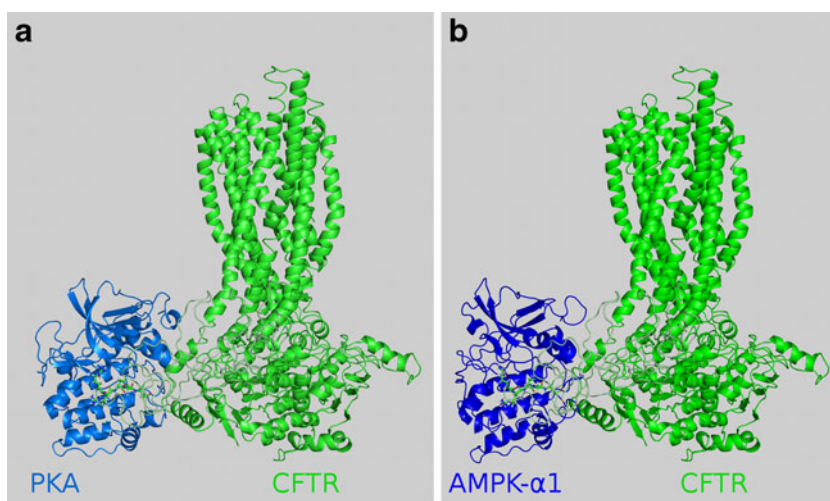
In our model, we observed the following H-bonds responsible for substrate peptide recognition between the recognized pentapeptide of CFTR (810–814) and AMPK, and between ADP and AMPK: CFTR Arg810–ADP; CFTR Arg811–AMPK Glu145, Tyr181; CFTR Ser813–AMPK

Lys142, Ser178; CFTR Gln814–AMPK Gly177. These bonds are presented in Fig. 2b. Lys142 is a residue in the active site associated with the kinase activity of AMPK. The overall structure of CFTR in complex with the AMPK catalytic domain is presented in Fig. 3b.

Comparison of the CFTR–AMPK catalytic domain and CFTR–PKA complexes

Our analysis revealed high probabilities of both complexes. The interface areas of both complexes measured with MSMS [32] differed slightly: 2548 \AA^2 for CFTR–PKA compared to 2149 \AA^2 for CFTR–AMPK, which suggests that the PKA–CFTR complex is more stable. However, both interface areas are typical of heterodimers. More detailed analysis with PIC [33] showed that although the AMPK–CFTR interface has only six hydrophobic interactions (in contrast to the nine present at the PKA–CFTR interface), and it does not form (main chain)–(main chain) hydrophobic bonds (in contrast to the three bonds formed in the PKA–CFTR case), its higher number of (main chain)–(side chain) hydrogen bonds (30 compared to 22) and (side chain)–(side chain) hydrogen bonds (29 compared to 20) suggest that both complexes have similar stabilities.

Fig. 3 Overall structures of the CFTR–PKA catalytic domain (*left*) and the CFTR–AMPK (*right*) complexes. Kinases and CFTR are represented in *cartoon form*, while the CFTR pentapeptide 810–814 is presented in both pictures in *stick form*



CFTR–AMPK regulatory domain complex

Experimental data for the CFTR–AMPK complex indicate that (i) the CFTR region 1420–1457 interacts with the 407–550 region of AMPK α 1 [16], and (ii) the R domain is phosphorylated by the catalytic domain of AMPK α 1 [12]. Prepared models of AMPK and CFTR lack unstructured tails: the final complex was defined as exposing terminal residues of the truncated subunits to the solvent in a way that enables chain extension. The AMPK regulatory domain was manually docked to CFTR in the experimentally confirmed interface area in a manner that conforms with established restrictions.

The CFTR interface region consists of two parts: fragment 1420–1428 is part of the NBD, while 1429–1450 is an unstructured CFTR tail. In our work, we modeled the interaction with a structured part (1420–1428), ensuring that the model provides the space required for the C-terminal tail. The complex was minimized. Interfaces did not introduce any energetically unfavorable conformations for any of the proteins; nor were any of the predicted H-bonds broken due to an unfavorable environment.

The structure of the CFTR–AMPK regulatory domain complex is shown in Fig. 4. In the interface created, 52 and 49 residues are involved on the CFTR and AMPK sides respectively, with a total area of 1856 Å². This is an average surface area for a nonobligatory complex interface. As further support for our model, we used information on mutations disrupting the interaction between AMPK α 1 and CFTR [16].

Discussion

Our models show that two counteracting kinases, PKA and AMPK, can phosphorylate CFTR S813, suggesting that the different effects of AMPK and PKA on CFTR Cl[−] channel activity result from interactions with other proteins and/or S813 availability.

AMPK binding to CFTR

Our model shows that AMPK regulatory binding to CFTR is not required for CFTR phosphorylation. However, the R domain undergoes constant structural changes, hampering the formation of an appropriate interface. Furthermore, the R domain has only two phospho sites in the precise AMPK-recognition motif (S737 and S768), which are inaccessible in the conformation used. The binding of AMPK to CFTR with the regulatory domain may provide the time required for the R domain to change conformation, thus allowing access to the “inhibitory” serines.

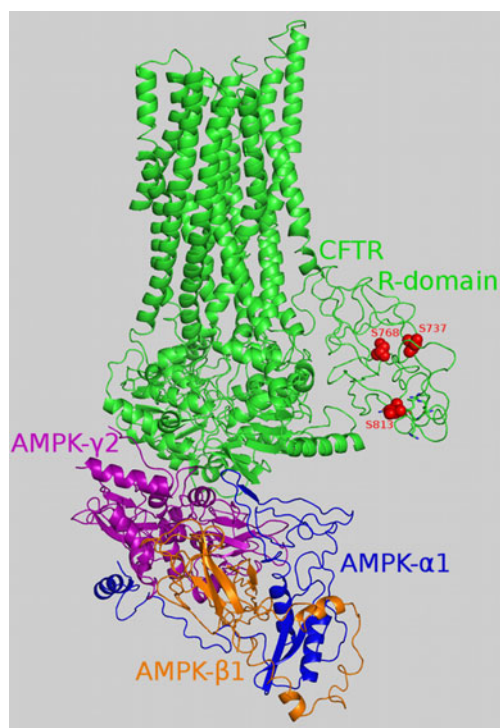


Fig. 4 Overall structure of the CFTR–AMPK regulatory domain complex. The main chains of the proteins are colored by chain, with the R domain highlighted in a lighter green color compared to the rest of the CFTR protein. Atoms from S737, S768 and S813 in the R domain are presented as red spheres. CFTR pentapeptide 810–814 is represented in stick form (except for S813, which is presented in sphere form)

S813 phosphorylation by PKA and AMPK

The complexes of CFTR with the catalytic subunit of PKA and the catalytic domain of AMPK α 1 are bound at the S813 site within the R domain.

Our model of S813 binding by PKA is consistent with experimental evidence: the active-site lysine (K169) is able to react with S813 and bind adenosine phosphate. AMPK bound to S813 in the same conformation as PKA, and did not differ from it in terms of the H-bonds formed and interface stability. This result suggests that AMPK may phosphorylate S813. This shows that, at the molecular level, it is possible that 20% of the currently unexplained phosphorylation signal observed in the CFTR double mutant S737A–S768A incubated with AMPK may come from “activator” serines such as S813.

Role of CFTR R-domain flexibility in phosphorylation

CFTR activity depends on R-domain phosphorylation [10, 40]. In the process of CFTR regulation, the R domain acts as an inhibitor that needs to be phosphorylated to enable CFTR channel activity [41]. Based on the results of

NMR experiments, Baker et al. described the R domain as “sampling multiple, heterogeneous conformations, [...] with rapid interconversion between conformers [5].” In our work, we decided to use one conformation among the many derived by Riordan et al. by de novo modeling [4]. In our model, the R domain’s structure prohibits the phosphorylation of the PKA and PKC phospho sites S686, S712, S768, and S790, which are known to be phosphorylated in vitro under controlled conditions. The latest experimental results show that the “inhibitory” serines S737 and S768 introduce an insensitive CFTR mutant into PKA (in which 15 potential phosphoserines are mutated to alanines) and restore PKA sensitivity [10]. Considering the diminishing differences between “activating” and “inhibiting” serines as well as “serines phosphorylated by activating and inhibiting kinases,” further structural analyses of the R domain and the nature of its interaction with suitable kinases appear to be crucial elements for elucidating CFTR regulation.

Our results present a molecular explanation for the phenomenon of the dependence of CFTR phosphorylation on the flexibility of the R domain. A new question can thus be formulated: what are the factors that influence the structural state of the R domain prior to phosphorylation?

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Wine JJ (1995) Cystic fibrosis: how do CFTR mutations cause cystic fibrosis? *Curr Biol* 5:1357–1359
2. Kirk KL (2000) New paradigms of CFTR chloride channel regulation. *Cell Mol Life Sci* 57:623–634
3. Ollero M, Brouillard F, Edelman A (2006) Cystic fibrosis enters the proteomics scene: New answers to old questions. *Proteomics* 14:4084–4099
4. Hegedűs T, Serohijos AWR, Dokholyan NV, He L, Riordan JR (2008) Computational studies reveal phosphorylation-dependent changes in the unstructured R domain of CFTR. *J Mol Biol* 378:1052–1063
5. Baker JMR, Hudson RP, Kanelis V, Choy W-Y, Thibodeau PH, Thomas PJ et al (2007) CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nat Struct Mol Biol* 14:738–745
6. Serohijos AWR, Hegedűs T, Aleksandrov AA, He L, Cui L, Dokholyan NV et al (2008) Phenylalanine-508 mediates a cytoplasmic–membrane domain contact in the CFTR 3D structure crucial to assembly and channel function. *Proc Natl Acad Sci USA* 105:3256–3261
7. Kongsuphol P, Cassidy D, Hieke B, Treharne KJ, Schreiber R, Mehta A et al (2009) Mechanistic insight into control of CFTR by AMPK. *J Biol Chem* 284:5645–5653
8. Treharne KJ, Xu Z, Chen JH, Best OG, Cassidy DM, Gruenert DC et al (2009) Inhibition of protein kinase CK2 closes the CFTR Cl⁻ channel, but has no effect on the cystic fibrosis mutant DeltaF508-CFTR. *Cell Physiol Biochem* 24:347–360
9. King JDJ, Fitch AC, Lee JK, McCane JE, Mak DO, Foskett JK et al (2009) AMP-activated protein kinase phosphorylation of the R domain inhibits PKA stimulation of CFTR. *Am J Physiol Cell Physiol* 297:C94–C101
10. Hegedűs T, Aleksandrov AA, Mengos A, Cui L, Jensen TJ, Riordan JR (2009) Role of individual R domain phosphorylation sites in CFTR regulation by protein kinase A. *Biochim Biophys Acta* 1788:1341–1349
11. Lewis HA, Buchanan SG, Burley SK, Connors K, Dickey M, Dorwart M et al (2004) Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. *EMBO J* 23:282–293
12. Kongsuphol P, Hieke B, Ousingsawat J, Almaca J, Viollet B, Schreiber R et al (2009) Regulation of Cl⁻ secretion by AMPK in vivo. *Pflugers Arch* 457:1071–1078
13. Carling D (2004) The AMP-activated protein kinase cascade: a unifying system for energy control. *Trends Biochem Sci* 29:18–24
14. Hallows KR, McCane JE, Kemp BE, Witters LA, Foskett KJ (2003) Regulation of channel gating by AMP-activated protein kinase modulates cystic fibrosis transmembrane conductance regulator activity in lung submucosal cells. *J Biol Chem* 278:998–1004
15. Mehta A (2005) CFTR: more than just a chloride channel. *Pediatr Pulmonol* 39:292–298
16. Hallows KR, Raghuram V, Kemp BE, Witters LA, Foskett KJ (2000) Inhibition of cystic fibrosis transmembrane conductance regulator by novel interaction with the metabolic sensor AMP-activated protein kinase. *J Clin Invest* 105:1711–1721
17. UniProt Consortium (2008) The universal protein resource. *Nucleic Acids Res* 36:D190–D195
18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H et al (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242
19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
20. Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244–W248
21. Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 48:443–453
22. Trafny EA, Xuong NH, Adams JA, Eyck LFT, Taylor SS, Sowadski JM, Madhusudan EA (1994) cAMP-dependent protein kinase: crystallographic insights into substrate recognition and phosphotransfer. *Protein Sci* 3:176–187
23. Townley R, Shapiro L (2007) Crystal structures of the adenylate sensor from fission yeast AMP-activated protein kinase. *Science* 315:1726–1729
24. Amodeo GA, Rudolph MJ, Tong L (2007) Crystal structure of the heterotrimer core of *Saccharomyces cerevisiae* AMPK homologue SNF1. *Nature* 449:492–495
25. Shindyalov IN, Bourne PE (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* 11:739–747
26. Ye Y, Godzik A (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* 19(Suppl 2):246–255
27. Lin X, Murray JM, Rico MX, Wang DT, Chu Y, Zhou M et al (2006) Discovery of 2-pyrimidyl-5-amidothiophenes as potent

- inhibitors for AKT: synthesis and SAR studies. *Bioorg Med Chem Lett* 15:4163–4168
28. Case DA, Pearlman DA, Caldwell JW, Cheatham TE, Ross WS, Simmerling CL et al (1999) AMBER6. *J Am Chem Soc* 117:5179–5197
 29. Schrödinger LLC (2010) PyMOL. <http://www.pymol.org>
 30. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 36:W197–W201
 31. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics* 22:195–201
 32. Sanner M, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 38:305–320
 33. Tina KG, Bhadra R, Srinivasan N (2007) PIC: Protein Interactions Calculator. *Nucleic Acids Res* 35:W473–W476
 34. Jin X, Townley R, Shapiro L (2007) Structural insight into AMPK regulation: ADP comes into play. *Structure* 15:1285–1295
 35. Oliver AW, Paul A, Boxall KJ, Barrie ES, Aherne WG, Garret MD et al (2006) Trans-activation of the DNA-damage signalling protein kinase Chk2 by T-loop exchange. *EMBO J* 25:3179–3190
 36. Chang X-B, Tabcharanie JA, Hou Y-X, Jensen TJ, Kartner N, Alon N et al (1993) Protein kinase A (PKA) still activates CFTR chloride channel after mutagenesis of all 10 PKA consensus phosphorylation sites. *J Biol Chem* 268:11304–11311
 37. Mio K, Ogura T, Mio M, Shimizu H, Hwang T-C, Sato C et al (2008) Three-dimensional reconstruction of human cystic fibrosis transmembrane conductance regulator chloride channel revealed an ellipsoidal structure with orifices beneath the putative transmembrane domain. *J Biol Chem* 283:30300–30310
 38. Wilkinson DJ, Strong TV, Mansoura MK, Wood DL, Smith SS, Collins FS et al (1997) CFTR activation: additive effects of stimulatory and inhibitory phosphorylation sites in the R domain. *Am J Physiol* 273:L127–L133
 39. Vais H, Zhang R, Reenstra WW (2004) Dibasic phosphorylation sites in the R domain of CFTR have stimulatory and inhibitory effects on channel activation. *Am J Physiol* 287:C737–C745
 40. Gadsby DC, Nairn AC (1999) Control of CFTR channel gating by phosphorylation and nucleotide hydrolysis. *Physiol Rev* 79:S77–S107
 41. Lewarchik CM, Peters KW, Qi J, Frizzel RA (2008) Regulation of CFTR trafficking by its R domain. *J Biol Chem* 283:28401–28412

Molecular electrostatic potentials of DNA base–base pairing and mispairing

Ivonne Otero-Navas · Jorge M. Seminario

Received: 24 December 2010 / Accepted: 14 February 2011 / Published online: 6 April 2011
© Springer-Verlag 2011

Abstract An understanding of why adenine (A) pairs with thymine (T) and cytosine (C) with guanine (G) in DNA is very useful in the design of sensors and other related devices. We report the use of dissociation energies, geometries and molecular electrostatic potentials (MEPs) to justify the canonical (AT and CG) Watson-Crick pairs. We also analyze all mismatches in both configurations—*cis* and *trans*—with respect to their glycoside bonds. As expected, we found that the most stable pair configuration corresponds to CG, providing an energy criterion for that preferred configuration. The reason why A gets together with T is much more difficult to explain as the energy of this pair is smaller than the energy of some other mismatched pairs. We tested MEPs to see if they could shed light on this problem. Interestingly, MEPs yield a unique pattern (shape) for the two canonical cases but different shapes for the mismatches. A tunnel of positive potential surrounded by a negative one is found interconnecting the three H-bonds of CG and the two of AT. This MEP tunnel, assisted partially by energetics and geometrical criteria, unambiguously determine a distinctive feature of the affinity between A and T as well as that between G and C.

Keywords Molecular electrostatic potential · Base pairing · Base mispairing · DNA energetics · Pair dissociation energy · Watson-Crick pair · Potential tunnel

Introduction

As the database of all organisms' biological information, the DNA molecule controls and coordinates all activities and cellular functions. Its molecular structure is a double helix formed by two strands that contains blocks of nucleotides; each nucleotide includes three parts: a phosphate group, a sugar group (deoxyribose sugar) and one of four bases: adenine (A), guanine (G), thymine (T) or cytosine (C) [1]. These bases are paired in a specific way: adenine always bonds with thymine and cytosine bonds with guanine (AT and CG, respectively). However, pairing errors (nucleobases occurring in non-complementary pairs) can occur within the DNA helix [2–4] and have been related directly to mutations [5] and to the development of cancer [6] and many other diseases [7].

Understanding the complexity of recognition and pairing mechanisms in DNA has been a major area of research [8–10]. Fareed Aboul-ela et al. [11], among others, have investigated the thermodynamic stability of mispaired bases with respect to canonical base pairs (AT and CG), finding that DNA molecules containing mismatches are destabilized. However, other studies [12] showed that damaged base pairs have structures that are very analogous to the Watson–Crick base pairs; for instance, AG [13] and GT [14] do not modify the DNA structure significantly and are among the most stable mismatches [11, 14–16]. Goodman [17] explained the selectivity of base pairing through size and hydrogen-bond complementarity, but some authors believe that the argument about distances is simply a consequence and not an intrinsic reason for pairing [18].

In addition, the central function of polymerase is the synthesis of a polynucleotide sequence from a nucleotide template. Recent studies have shown that a suitable way to

I. Otero-Navas · J. M. Seminario (✉)
Department of Chemical Engineering, Department of Electrical
and Computer Engineering, Texas A&M University,
College Station,
Texas, USA
e-mail: seminario@tamu.edu

understand base–base pairing is to evaluate the mechanism by which polymerase chooses the base complements [19–21]; however, these investigations reached contradictory results. Some reported that hydrogen bonds do not play a substantial role in selective pairing [22–24]; while others demonstrated new examples of hydrogen bonded pairs in which the H-bonding arrangement is different from the canonical AT and GC [25, 26]. Other investigations showed several mismatches produced by polymerases, depending on their enzyme types [19, 22]. For instance, Moore et al. found that human DNA primase processes DNA bases poorly, and implied the importance of hydrogen bonds in pairing [19].

Theoretical studies that evaluated molecular interactions energies and hydrogen bonding in Watson-Crick base pairs found that the geometry, bond lengths, polarization, charge transfer and dispersion interactions have a considerable effect on the attraction energy of bases [27–29]. Other authors have studied the canonical (AT and CG) and non-canonical (all pair combinations of A, T, C, and G except for AT and CG) base pairs in RNA because all these pairs have strong structural and functional implications for the catalytic or regulatory functions of genetic information [30, 31].

In this work, we present a theoretical study of the molecular electrostatic potentials (MEPs), adiabatic dissociation energy, vertical dissociation energy, and other characteristics such as bond lengths to explain how the bases recognize each other leading to canonical base-base pairing, including some of the reasons for which mismatches occur in DNA. The following sections explain our [Methods](#) and [Results](#).

Methods

The geometries of canonical and mismatched base pairs were optimized until local minima were obtained using ab initio molecular orbital theory with the Gaussian-03 [32] and Gaussian-09 [33] programs. Mispairing geometries feature cis and trans conformations, depending on the relative positions of their glycoside bonds, i.e., those connecting to their sugar backbones, from the intersection to a line drawn through the H-bonds connecting the paired bases [34–36]. We added hydrogen atoms instead of sugars for the sake of valence bookkeeping. We use the Becke three-parameter hybrid exchange functional [37] with the generalized gradient approximation correlation functional Perdew-Wang-91 [38] and the basis set 6-31 G(*d,p*), both of which include polarization functions in all atoms [39, 40]. Locations of all atoms were optimized searching for zero forces (first derivatives) in all atoms. A second derivative calculation was performed in order to verify that the

Fig. 1 Optimized geometries for paired (*green frame*) and mispaired (*blue frame*) bases. H bond lengths (*a, b, c*) in angstroms (Å), carbon (*dark gray*), oxygen (*red*), nitrogen (*blue*), and hydrogen (*light gray*)

optimized geometries do not show negative eigenvalues in the Hessian matrices, verifying that all structures correspond to local minima. All these procedures have been tested widely in the past with energetic materials [41–47], molecular devices [48–50], metallic clusters [51–54], and combined procedures of molecular dynamics and density functional theory (DFT) [43, 55] calculations. DFT improves over HF-based calculations, which provides good qualitative results [56, 57]. The MEPs, $V(\vec{r})$, that nuclei and electrons induce at each point r of space can be calculated from [58–61],

$$V(\vec{r}) = \sum_i \frac{Z_i e}{|\vec{R}_i - \vec{r}|} - \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} dr' \quad (1)$$

after solving the Schrödinger equation and obtaining ρ from the wavefunction. Z_i is the atomic number of atom i situated at \vec{R}_i ; ρ is the electron density and the integral is over the whole space. ρ can be calculated from the wavefunction obtained solving the Schrödinger equation. This could be done using DFT [62–64] or conventional wavefunction methods [65–67]. The reader is referred to reference [65] for further details on calculations using the Hartree-Fock (HF) procedure to determine the MEPs of the nucleotide bases. In the present work, the MEPs are calculated and visualized on planes and closed surfaces from the optimized canonical and mismatched base pairs.

We also calculated the adiabatic and vertical dissociation energies, which were calculated using

$$\Delta E_a(ij) = E(i) + E(j) - E(ij) \quad (2)$$

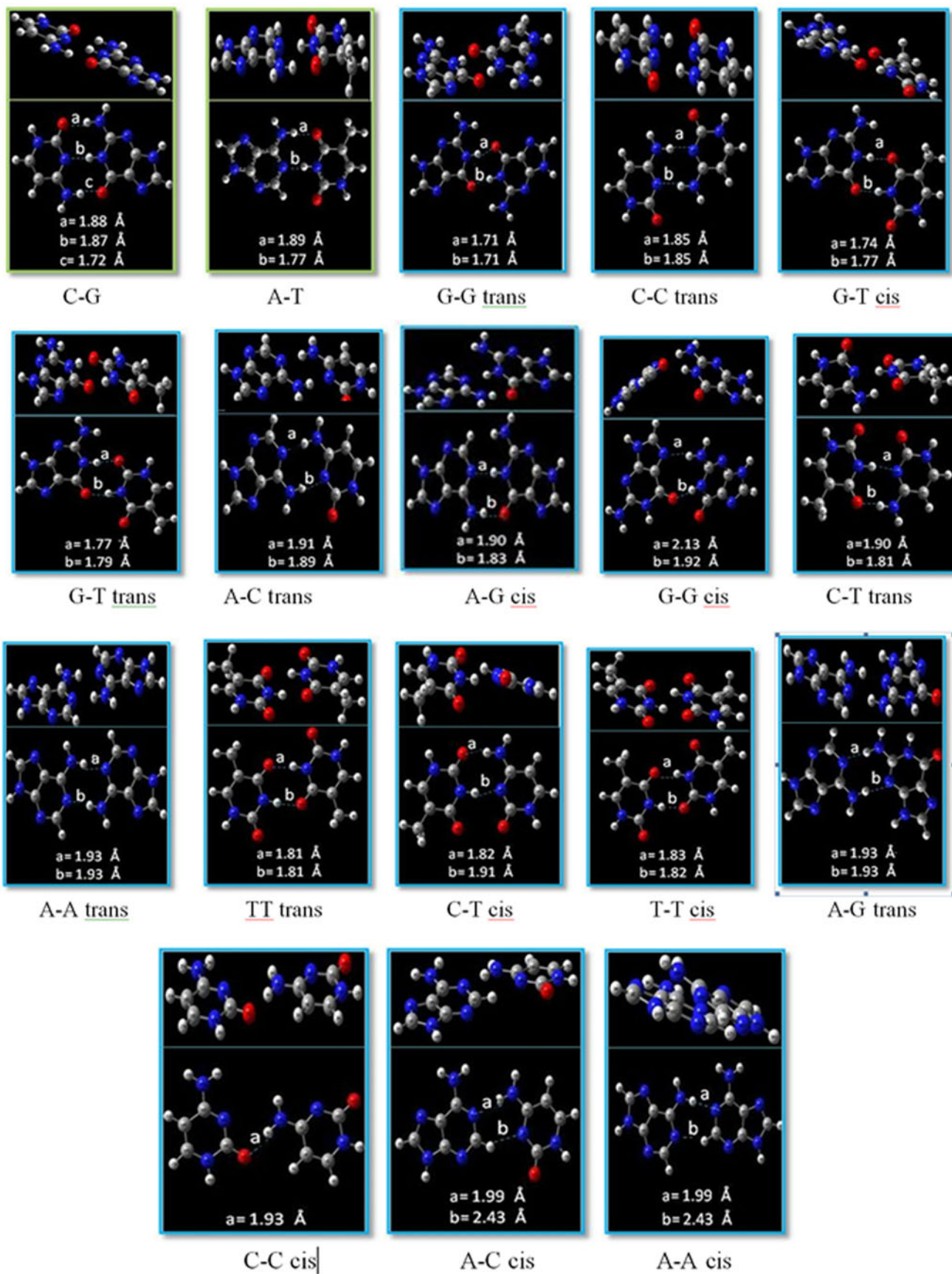
and

$$\Delta E'_v(ij) = E'(i) + E'(j) - E'(ij), \quad (3)$$

respectively, where $E(i)$ and $E(j)$ are the ground state energies of any of the four optimized nucleotides making the pair ij , and $E(ij)$ is the energy corresponding to the optimized canonical and mismatched base pairs. $E'(i)$ and $E'(j)$ are the energies of the individual pairs obtained exactly with the geometries they have in the complex pair (i.e., without any structural relaxation of the i and j nucleotides).

Results

Figure 1 shows the optimized structures for base–base pairing and mispairings. We did not consider solvent and ion interactions as we did in [68] because of the small effect



of those interactions on the MEPs in the H-bonded regions of the paired nucleotides. The solvent and other environmental factors may affect DNA under normal working conditions [69–72] but these effects on the MEPs in the regions of interest for our study, i.e., in between bases connected by H-bonds, are very limited. This region is anyway a special case of the paired and mispaired DNA bases as solvent molecules do not enter this region.

Figure 2 also shows the nonbonded part of the H-bond distances. The shortest distances in the CG pair are 1.72 and 1.77 Å in the AT. The shortest lengths of the mismatch vary from 1.71 to 1.99. Thus, several mismatches fell within the normal values of the canonical pairs, a few others are up to 0.22 Å. Since these distances are strongly adaptable to the environment, we do not consider that the actual length between H-bonds is a factor that determines base–base pairing. In addition, with the exception of AC-cis and AA-cis, with distances of 1.99 Å, all the others are no longer than 1.93 Å, i.e., no more than 0.16 Å longer than the AT case. On the other hand, a strong geometrical factor related to the length of the pair is the fact that there are two bases that are really much shorter than the other two. C and T are made of only one six-membered ring but A and G are made of a five-member ring fused to a six-membered ring; thus the two-ring bases are more than 1 Å longer than the one-ring base. We can conclude that the distance factor can serve to rule out pairs such as CT, CC, TT (short pairs) and AG, AA, GG (long pairs), leaving as possible pairs the combinations AT, CG, AC, and GT, which connect a one-ring to a two-ring base through H-bonds.

Table 1 shows the energies at the optimized geometries for the four bases; these were used to obtain the adiabatic binding energies of the pairs shown in Table 2, which also shows the dissociation energies for canonical and non-canonical Watson-Crick pairs. Mismatches in Table 2 are organized according to their adiabatic energies. The CG base pair has the highest vertical and dissociation energies; thus, this structure has the strongest interaction between pairs. The structure of CG is unique in that it has three hydrogen bonds, two NH–O, and one NH–N type, while AT has a lower dissociation energy (15.3 kcal mol⁻¹) than CG. This result is in agreement with the fact that the stability of the DNA double helix is affected by the CG content. For this reason, DNA sequences rich in CG are more stable structures than AT-rich sequences [73]. The interaction between guanine and cytosine is mediated by electrostatically driven hydrogen bonding between atoms that are positively (hydrogen atoms) and negatively (nitrogen and oxygen atoms) polarized. This yields a CG structure with the most favorable energy. In contrast, AT has an adiabatic dissociation energy smaller than those of CG, GG-trans, CC-trans, GT-cis, GT-trans and AC-trans;

therefore, the energetics for AT cannot explain the selectivity between adenine and thymine. The structures with the lowest adiabatic and vertical dissociation energy are AC-cis and AA-cis; they both have CH–N and NH–N mediated hydrogen bonds, while the other pairs are stabilized by N–HN and O–HN hydrogen bonds. This indicates that pairs involving strong N–HN and NH–O hydrogen bonds have larger conformational stability. Consistent with this, Fig. 1 shows AC-cis and AA-cis as the pairs with the longest bond lengths.

Both AT and CG have planar conformations (Fig. 1), but mismatches such as GG-trans, CC-trans, GT-cis, GT-trans, AC-trans, AA-trans, TT-trans, AG-trans and TT-cis also have planar conformations (Fig. 1); this is another argument to indicate that geometrical arrangement alone does not describe base pairing completely. These results are consistent with previous investigations by other authors, where incorporation of some mismatches causes geometrical distortion in a small DNA helix [12, 15, 74]. However, structures with lower adiabatic and vertical dissociation energies than AT, such as AG-cis, GG-cis, CT-trans, CT-cis, AC-cis and AA-cis, have the largest deviations from a planar conformation.

MEPs for each of the paired and mispaired bases show regions with high dominance of electrons over nuclei because of oxygen and nitrogen lone pairs, and regions with low electron dominance localized on hydrogen atoms. The canonical and non-canonical Watson-Crick conformations are driven by electrostatic interactions between zones with negative and positive potentials (Fig. 2c), suggesting that conformations other than AT and CG might occur. For instance, in GT-cis, one of the hydrogen bonds is formed between H (Guanine) with positive potential and O (thymine) with negative potential.

However, we find that canonical and non-canonical pairs exhibit distinct MEP surfaces (Fig. 2d) on a plane at the center of the nonbonded part of the H-bonds, thus suggesting that $V(r)$ can be used to distinguish between base-base pairing and mispairing. Analyzing MEPs on a central plane between the bases (Fig. 2e), we find a common pattern in canonical pairing: CG and AT shows a positive potential tunnel surrounded by a negative one; this pattern differs from those of the mismatches, where no tunnel is found. Even the energetically favorable structures, with adiabatic dissociation energies higher than that of the AT pair, show distortions in the potential tunnel. These differences in MEPs are due to the chemical structure of the pairs, as reflected in the electron and nuclei in the neighborhood of the atoms involved in hydrogen bonding. The exception to the distorted tunnel for non-canonical pairs are CC-trans, TT-cis, and TT-trans, which yield MEPs planes with shapes very similar to the canonical tunnel. These three pairs could be correlated with deficiencies in

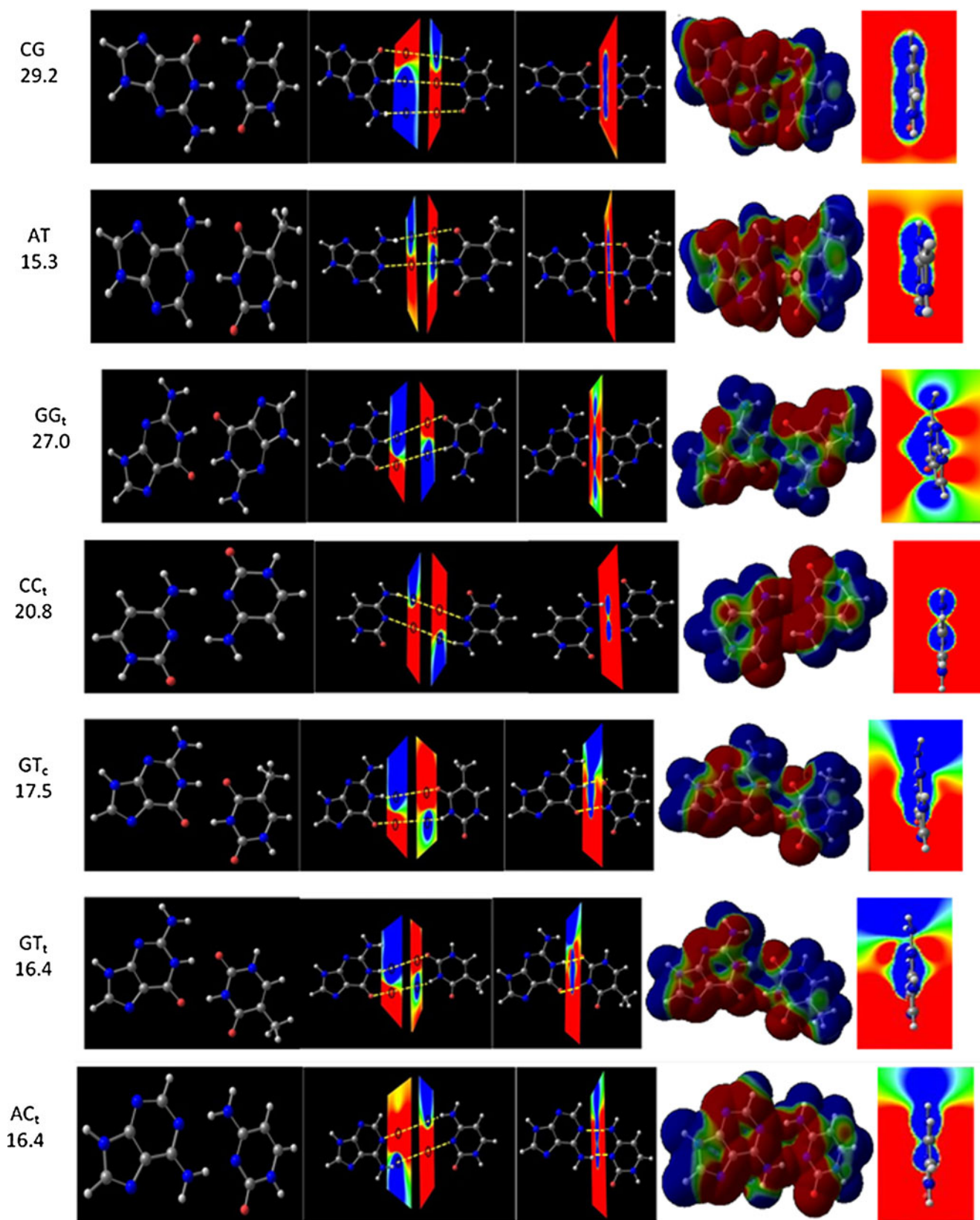


Fig. 2 **a–e** Results of base-base pairing and mispairing analyses. **a** Pair type and H-bond dissociation energy (kcal mol^{-1}), subscripts c=cis and t=trans. **b** Optimized geometries for canonical and non-canonical Watson-Crick base pairs. **c** Molecular electrostatic potential (MEP)

planes located in the neighborhood of each base. **d** MEP plane in the middle of hydrogen bond. **e** MEP surface for canonical and non-canonical Watson-Crick base pairs. **f** frontal view of MEP plane shown in **d**

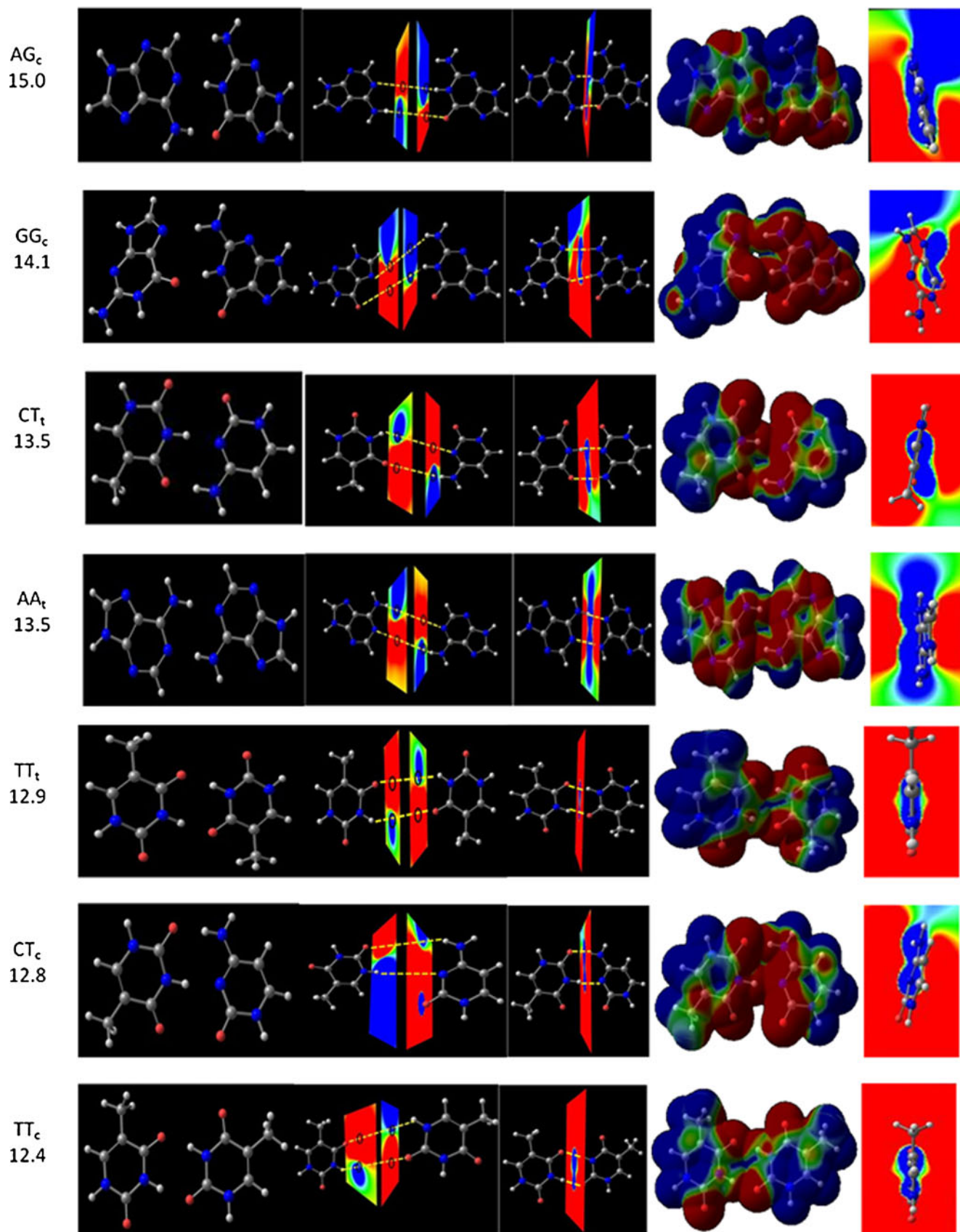


Fig. 2 (continued)

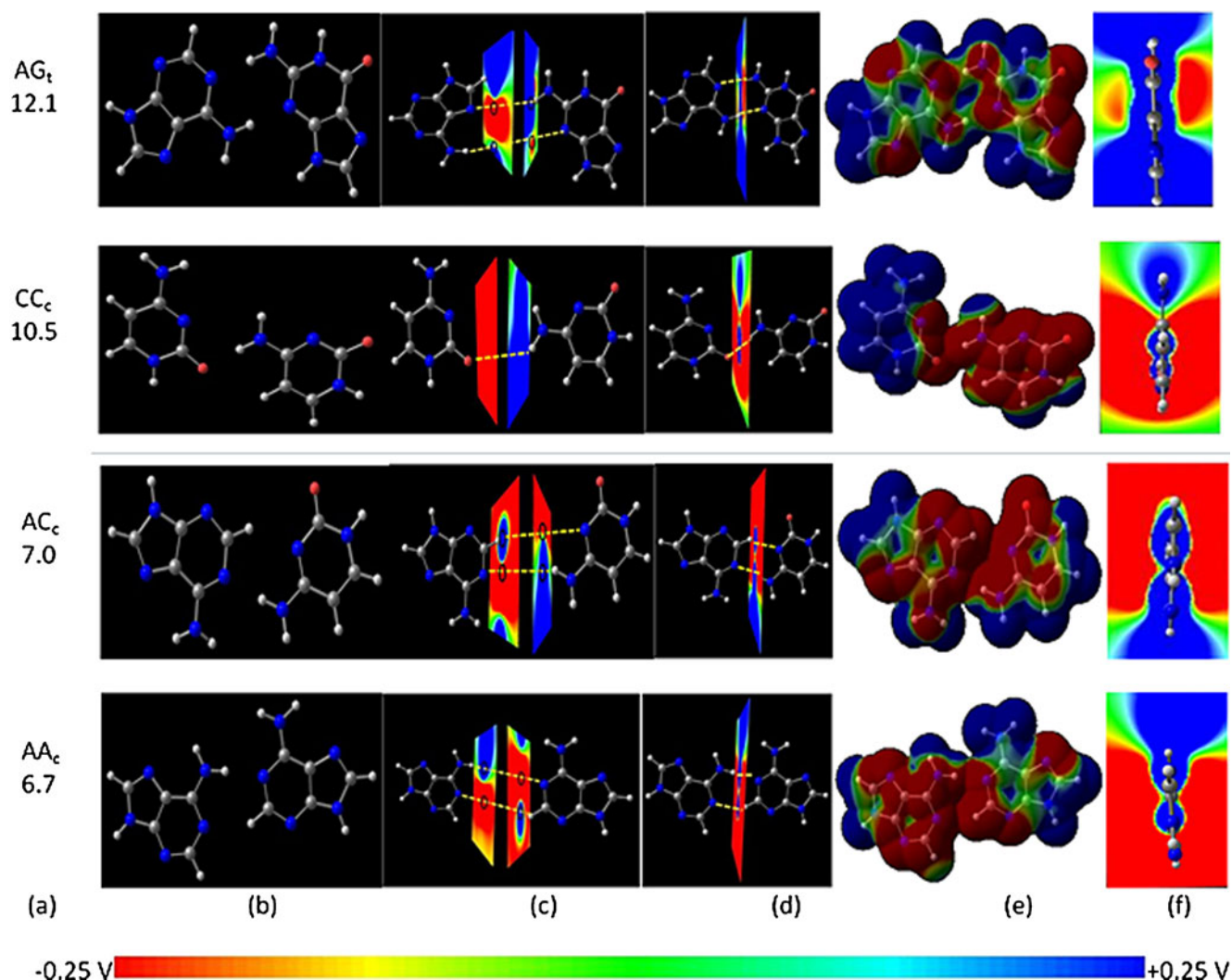


Fig. 2 (continued)

DNA mismatch recognition mechanisms.; thus, MEP shape may also be an indicator of mismatch probability. Nevertheless, the above non-canonical cases can be dismissed because all three correspond to short length pairs. It has also been shown that these three mispaired bases are corrected by the mismatch repair system with low efficiency, perhaps because of the close similarity between their MEPs and those of the canonical MEPs. Meanwhile, other mispaired bases such as GT, AC, GG, AG and AA are repaired with high and intermediate effectiveness [11, 14, 75]; these mispaired bases have MEPs planes with positive regions in sites where the canonical base pairs have a strong negative potential, thus, we suggest that these mispaired bases are easily distinguishable by nature's correcting mechanisms. Therefore, we can relate the pairing mechanism in DNA with $V(r)$; thus enzymes entrusted to duplication processes might use MEP patterns to perform the pairing or the mismatch repair.

Conclusions

We analyzed MEPs, geometries and energetics to explain canonical Watson-Crick base pairs, and found that the CG pair has the highest dissociation energy, having a unique structure of three hydrogen bonds; however, the AT pair does not possess the energetics criterion for selectivity. We find that other mismatches (CC- and GG-trans, CC-trans, GT-cis, GT-trans and AC-trans) have adiabatic and vertical dissociation energies higher than AT. On the other hand, the length of the

Table 1 Total energies (E) of the four DNA bases at the B3PW91/6-31 G(d,p) level of theory

Base	E (Ha)
A	-467.15765
T	-453.97828
C	-394.79299
G	-542.36450

Table 2 Total energy (E_{ij}), adiabatic dissociation energy (ΔE_a), vertical dissociation energies (ΔE_v) basis set superposition error (BSSE), zero point energy (ZPE), and H-bond lengths d for base-base paired and mispaired, all at the B3PW91/6-31 G(d,p) level of theory

ij	E_{ij} (Ha)	ΔE_a (kcal/mol)	ΔE_v (kcal/mol)	BSSE (kcal/mol)	ZPE (kcal/mol)	d (Å)	Donor lone pair	Acceptor
GC	-937.20405	29.2	33.4	3.9	1.8	O-HN=1.88	O	HN
						NH-N=1.87	N	HN
						NH-O=1.72	O	HN
AT	-921.16034	15.3	17.2	3.4	1.2	NH-O=1.89	O	HN
						N-HN=1.77	N	HN
GGt	-1084.77196	27.0	32.2	3.9	0.8	NH-O=1.71	O	HN
						O-HN=1.71	O	HN
CCt	-789.61919	20.8	23.4	2.5	2.2	NH-N=1.85	N	HN
						N-HN=1.85	N	HN
GTc	-996.37061	17.5	20.2	3.3	0.8	NH-O=1.74	O	HN
						O-HN=1.77	O	HN
GTt	-996.36887	16.4	16.9	3.2	0.7	NH-O=1.77	O	HN
						O-NH=1.79	O	HN
ACt	-861.97674	16.4	18.9	2.4	1.8	N-HN=1.91	N	HN
						NH-N=1.89	N	HN
AGc	-1009.54605	15.0	18.2	3.1	1.6	N-HN=1.90	N	HN
						NH-O=1.83	O	HN
GGc	-1084.75152	14.1	14.8	2.1	1.1	N-HN=2.13	N	HN
						O-HN=1.92	O	HN
CTt	-848.79281	13.5	15.7	2.9	1.3	NH-N=1.90	N	HN
						O-HN=1.81	O	HN
AAt	-934.33680	13.5	14.9	2.3	1.6	NH-N=1.93	N	HN
						N-HN=1.93	N	HN
TTt	-907.97714	12.9	14.4	3.0	0.7	O-HN=1.81	O	HN
						NH-O=1.81	O	HN
CTc	-848.79159	12.8	14.9	2.7	1.2	O-HN=1.82	O	HN
						NH-N=1.91	N	HN
TTc	-907.97633	12.4	13.8	2.9	0.8	O-HN=1.83	O	HN
						NH-O=1.82	O	HN
AGt	-1009.54143	12.1	13.9	2.4	0.9	N-HN=1.94	N	HN
						NH-N=1.98	N	HN
CCc	-789.60268	10.5	11.1	1.6	1.4	O-HN=1.93	O	HN
ACc	-861.96185	7.0	7.8	1.9	1.3	N-HN=1.99	N	HN
						CH-N=2.43	N	HC
AAc	-934.32594	6.7	7.2	1.9	1.2	NH-N=2.01	N	HN
						N-HC=2.47	N	HC

individual bases may yield too short and too long paired bases that can be discarded from the list of possible base–base pairs. However, we note that paired and mispaired bases show differences in MEP surfaces; the canonical pairs show a common pattern in the MEPs located at the central plane between bases: a positive potential region, like a tunnel from base to base, is surrounded by a negative potential. We conclude that this shape, along with some information from the energetics and geometry criteria, explain unambiguously the selectivity between bases in DNA. Hence, pairing and repair

mechanisms are related to MEPs, which allows us, and perhaps nature, to discriminate the canonical paired bases from the mispaired ones.

Although the thermodynamic stability of canonical Watson-Crick base pairs has been the subject of literally hundreds of papers from the standpoint of biophysics and computational chemistry, it could be argued that the use of computational methods requires sophisticated methods and basis sets; nevertheless, it is possible to evaluate the base pairs from the electronic point of view using the MEPs,

which are not affected strongly by the double-helix or by taking into consideration the competitive hydrogen bonding with the solvent. The molecular potentials are intrinsic properties of the molecules that are affected by nonbonded interactions. Whatever takes place outside of the H-bonds between a pair does not affect strongly the results obtained in this work, including the ignoring of finite temperatures. MEPs are practically independent of changes in temperature below or around 300 K.

In practical terms, the computations we perform ensure that errors do not affect the MEPs. The B3PW91/6-31 G(d) level of theory has been tested widely to yield excellent results in several aspects of chemistry, including energetics. On the other hand, the use of a suitable basis set, the size of the molecules, and the level of exchange and correlation used in the B3PW91 hybrid functional also ensure that the MEPs calculated in this work do not suffer from basis set superposition errors (BSSE); corrections affect all complex pairs roughly to the same extent, perhaps because of the very similar type of neighboring atoms. Other errors, for instance, errors due to the lack of dispersion interactions practically do not affect the shape of the calculated MEPs for the H-bonded pairs. Zero-point energy (ZPE) relative corrections for all pairs are within 0.8 and 1.8 kcal mol⁻¹ thus their effect at most is 1 kcal mol⁻¹, which does not affect the results and has no effect on the MEP. Notice that we are using the qualitative features of MEPs rather than quantitative features. It is worth emphasizing that the MEP reflects the combined effects of both the nuclei and the electrons. In this sense, the MEP corresponds to another very useful function for describing features of molecular systems in addition to wavefunctions and density functions; however, MEP is by far the most practical function to use in such descriptions.

In summary, we have found an unambiguous way to determine the conditions for pairing between the four bases of DNA, which will be extremely useful for applications that mimic nature, such as in sensors, rapid DNA sequencing, and molecular recognition.

Acknowledgments We thank the Defense Threat Reduction Agency (DTRA) through the Army Research Office (ARO) for their support, Project No. W911NF-06-1-0231 and the ARO for the DURIP/ARO Project # W911NF-07-1-0199 and the MURI/ARO Project # W911NF-11-1-0024.

References

- Watson JD, Crick FHC (1953) Molecular structure of nucleic acids—a structure for deoxyribose nucleic acid. *Nature* 171:737–738
- Kornberg A (1980) DNA replication. Freeman, San Francisco
- Nir E, Plützer C, Kleinermanns K, de Vries M (2002) Properties of isolated DNA bases, base pairs and nucleosides examined by laser spectroscopy. *Eur Phys J D* 20:317–329
- Showalter AK, Tsai MD (2001) A DNA polymerase with specificity for five base pairs. *J Am Chem Soc* 123:1776–1777. doi:10.1021/ja005758x
- Friedberg EC, Walker GC, Siebe W (1995) DNA repair and mutagenesis. ASM, Washington, DC
- Loft S, Poulsen HE (1996) Cancer risk and oxidative DNA damage in man. *J Mol Spectrosc Med* 74:297–312
- Nir E, Kleinermanns K, de Vries MS (2000) Pairing of isolated nucleic-acid bases in the absence of the DNA backbone. *Nature* 408:949–951
- Krueger AT, Kool ET (2007) Model systems for understanding DNA base pairing. *Curr Opin Chem Biol* 11:588–594
- Edirisinghe N, Apalkov V, Berashevich J, Chakraborty T (2010) Electrical current through DNA containing mismatched base pairs. *Nanotechnology* 21:245101
- Apalkov V, Berashevich J, Chakraborty T (2010) Unique magnetic signatures of mismatched base pairs in DNA. *J Chem Phys* 132:085102
- Aboul-ela F, Koh D, Tinoco IJ (1985) Base-base mismatches. Thermodynamics of double helix formation for dCA3XA3G+dCT3YT3G (X, Y=A, C, G, T). *Nucleic Acids Res* 13:4811–4824
- Schofield MJ, Hsieh P (2003) DNA MISMATCH REPAIR: Molecular mechanisms and biological function. *Annu Rev Microbiol* 57:579–608. doi:10.1146/annurev.micro.57.030502.090847
- Modrich P (1987) DNA Mismatch correction. *Annu Rev Biochem* 56:435–466. doi:10.1146/annurev.bi.56.070187.002251
- Werntges H, Steger G, Riesner D, Fritz HJ (1986) Mismatches in DNA double strands: thermodynamic parameters and their correlation to repair efficiencies. *Nucleic Acids Res* 14:3773–3790. doi:10.1093/nar/14.9.3773
- Marra G, Schär P (1999) Recognition of DNA alterations by the mismatch repair system. *Biochem J* 338:1–13
- Patel DJ, Ikuta S, Kozlowski S, Itakura K (1983) Sequence dependence of hydrogen exchange kinetics in DNA duplexes at the individual base pair level in solution. *Proc Natl Acad Sci USA* 80:2184–2188
- Goodman MF (1997) Hydrogen bonding revisited: Geometric selection as a principal determinant of DNA replication fidelity. *Proc Natl Acad Sci USA* 94:10493–10495
- Jauregui LA, Seminario JM (2008) A DNA sensor for sequencing and mismatches based on electron transport through Watson–Crick and Non-Watson–Crick base pairs. *IEEE Sensors J* 8:803–814
- Moore CL, Zivkovic A, Engels JW, Kuchta RD (2004) Human DNA primase uses Watson–Crick hydrogen bonds to distinguish between correct and incorrect nucleoside triphosphates. *Biochemistry* 43:12367–12374. doi:10.1021/bi0490791
- Wolfe WT, Washington MT, Kool ET, Spratt TE, Helquist SA, Prakash L, Prakash S (2005) Evidence for a Watson–Crick hydrogen bonding requirement in DNA synthesis by human DNA polymerase {kappa}. *Mol Cell Biol* 25:7137–7143. doi:10.1128/mcb.25.16.7137-7143.2005
- Mizukami S, Kim TW, Helquist SA, Kool ET (2006) Varying DNA base-pair size in subangstrom increments: evidence for a loose, not large, active site in low-fidelity Dpo4 polymerase. *Biochemistry* 45:2772–2778. doi:10.1021/bi051961z
- Kim TW, Delaney JC, Essigmann JM, Kool ET (2005) Probing the active site tightness of DNA polymerase in subangstrom increments. *Proc Natl Acad Sci USA* 102:15803–15808. doi:10.1073/pnas.0505113102
- Kim TW, Briebe LG, Ellenberger T, Kool ET (2006) Functional evidence for a small and rigid active site in a high fidelity DNA polymerase. *J Biol Chem* 281:2289–2295. doi:10.1074/jbc.M510744200
- Mitsui T, Kimoto M, Harada Y, Yokoyama S, Hirao I (2005) An efficient unnatural base pair for a base-pair-expanded transcription system. *J Am Chem Soc* 127:8652–8658. doi:10.1021/ja0425280

25. Yang Z, Sismour AM, Sheng P, Puskar NL, Benner SA (2007) Enzymatic incorporation of a third nucleobase pair. *Nucleic Acids Res* 35:4238–4249. doi:10.1093/nar/gkm395
26. Sismour AM, Benner SA (2005) The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. *Nucleic Acids Res* 33:5640–5646. doi:10.1093/nar/gki873
27. Danilov VI, Anisimov VM (2005) Post Hartree-Fock studies of the canonical Watson-Crick DNA base pairs: molecular structure and the nature of stability. *J Biomol Struct Dyn* 22:471–482
28. Hobza P, Sponer J (1999) Structure, energetics, and dynamics of the nucleic acid base pairs: nonempirical ab initio calculations. *Chem Rev* 99:3247–3276
29. Guerra CF, Bickelhaupt FM, Snijders JG, Baerends EJ (2000) Hydrogen bonding in DNA base pairs: reconciliation of theory and experiment. *J Am Chem Soc* 122:4117–4128
30. Das J, Mukherjee S, Mitra A, Bhattacharyya D (2006) Non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. *J Biomol Struct Dyn* 24:149–161
31. Bhattacharyya D, Koripella SC, Mitra A, Rajendran VB, Sinha B (2007) Theoretical analysis of noncanonical base pairing interactions in RNA molecules. *J Biosci* 32:809–825
32. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03, Revision C.02. Gaussian Inc, Wallingford CT
33. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian 09, Revision A.02. Gaussian Inc, Wallingford CT
34. Lescoute A, Westhof E (2006) The interaction networks of structured RNAs. *Nucleic Acids Res* 34:6587–6604
35. Leontis NB, Westhof E (1998) Conserved geometrical base-pairing patterns in RNA. *Q Rev Biophys* 31:399–455. doi:10.1017/S0033583599003479
36. Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* 7:499–512
37. Becke AD (1993) A new mixing of Hartree-Fock and local density-functional theories. *J Chem Phys* 98:1372–1377
38. Perdew JP (1991) Unified theory of exchange and correlation beyond the local density approximation. In: Ziesche P, Eschrig H (eds) *Electronic structure of solids*. Akademie, Berlin, pp 11–20
39. Petersson GA, Bennett A, Tensfeldt TG, Al-Laham MA, Shirley WA, Mantzaris J (1988) A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. *J Chem Phys* 89:2193–2218
40. Petersson GA, Al-Laham MA (1991) A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *J Chem Phys* 94:6081–6090
41. Habibollahzadeh D, Grodzicki M, Seminario JM, Politzer P (1991) Computational study of the concerted gas-phase triple dissociations of 1,3,5-triazacyclohexane and its 1,3,5-trinitro derivative (RDX). *J Phys Chem* 95:7699–7702
42. Politzer P, Seminario JM (1993) Computational study of the structure of dinitraminic acid, $\text{HN}(\text{NO}_2)_2$ and the energetics of some possible decomposition steps. *Chem Phys Lett* 216:348–352
43. Seminario JM, Concha MC, Politzer P (1995) A density functional/molecular dynamics of the structure of liquid nitromethane. *J Chem Phys* 102:8281–8282
44. Politzer P, Seminario JM (1993) Energy changes associated with some decomposition steps of 1,3,3-trinitroazetidine—a nonlocal density-functional study. *Chem Phys Lett* 207:27–30
45. Politzer P, Seminario JM, Bolduc PR (1989) A proposed interpretation of the destabilizing effect of hydroxyl-groups on nitroaromatic molecules. *Chem Phys Lett* 158:463–469
46. Seminario JM, Concha MC, Politzer P (1992) Calculated structures and relative stabilities of furoxan, some 1,2 dinitrosoethylenes and other isomers. *J Comput Chem* 13:177–182
47. Murray JS, Redfern PC, Seminario JM, Politzer P (1990) Anomalous energy effects in some aliphatic and alicyclic aza systems and their nitro-derivatives. *J Phys Chem A* 94:2320–2323
48. Seminario JM, Araujo RA, Yan L (2004) Negative differential resistance in metallic and semiconducting clusters. *J Phys Chem B* 108:6915–6918
49. Seminario JM, De La Cruz C, Derosa PA, Yan L (2004) Nanometer-size conducting and insulating molecular devices. *J Phys Chem B* 108:17879–17885
50. Derosa PA, Guda S, Seminario JM (2003) A programmable molecular diode driven by charge-induced conformational changes. *J Am Chem Soc* 125:14240–14241
51. Seminario JM, Zacarias AG, Castro M (1997) Systematic study of the lowest energy states of Pd, Pd₂, and Pd₃. *Int J Quantum Chem* 61:515–523
52. Seminario JM, Tour JM (1997) Systematic study of the lowest energy states of Au_n (n=1–4) Using DFT. *Int J Quantum Chem* 65:749–758
53. Derosa PA, Seminario JM, Balbuena PB (2001) Properties of small bimetallic Ni-Cu clusters. *J Phys Chem A* 105:7917–7925
54. Seminario JM, Ma Y, Agapito LA, Yan L, Araujo RA, Bingi S, Vadlamani NS, Chagarlamudi K, Sudarshan TS, Myrick ML, Colavita PE, Franzon PD, Nckashi DP, Cheng L, Yao Y, Tour JM (2004) Clustering effects on discontinuous gold film nanocells. *J Nanosci Nanotechnol* 4:907–917
55. Seminario JM, Derosa PA, Cordova LE, Bozard BH (2004) A molecular device operating at Terahertz frequencies. *IEEE Trans Nanotech* 3:215–218
56. Choi D-S, Huang S, Huang M, Barnard TS, Adams RD, Seminario JM, Tour JM (1998) Revised structures of N-substituted dibrominated pyrrole derivatives and their polymeric products. Termaleimide models with low optical bandgaps. *J Org Chem* 63:2646–2655
57. Murray JS, Seminario JM, Politzer P (1994) Does antiaromaticity imply net destabilization. *Int J Quantum Chem* 49:575–579
58. Politzer P, Seminario JM (1989) Computational analysis of the structures, bond properties, and electrostatic potentials of some nitrotetrahydrones and nitroazetatetrahydrones. *J Phys Chem* 93:4742–4745

59. Seminario JM, Politzer P (1991) First principles theoretical methods for the calculation of electronic charge densities and electrostatic potentials. In: Jeffrey GA (ed) The application of charge-density research to chemistry and drug design. Plenum, New York
60. Politzer P, Grice ME, Murray JS, Seminario JM (1993) Anomalous stabilizing and destabilizing effects in some cyclic pi-electron systems. *Can J Chem* 71:1123–1127
61. Seminario JM, Concha MC, Murray JS, Politzer P (1994) Theoretical analyses of O₂/H₂O systems under normal and supercritical conditions. *Chem Phys Lett* 222:25–32
62. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. *Phys Rev* 136:B864
63. Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. *Phys Rev* 140:A1133
64. Sham LJ, Kohn W (1966) One-particle properties of an inhomogeneous interacting electron gas. *Phys Rev* 145:561–567
65. Murray JS, Peralta-Inga Z, Politzer P, Ekanayake K, LeBreton P (2001) Computational characterization of nucleotide bases: Molecular surface electrostatic potentials and local ionization energies, and local polarization energies. *Int J Quantum Chem* 83:245–254
66. Politzer P, Seminario JM (1990) Relative bond strengths in tetrahedrane, prismane, and some of their aza analogs. *Struct Chem* 1:29–32
67. Scrocco E, Tomasi J (1973) The electrostatic molecular potential as a tool for the interpretation of molecular properties. In: New concepts II, vol 42. Topics in current chemistry. Springer, Berlin, pp 95–170. doi:10.1007/3-540-06399-4_6
68. Jauregui LA, Salazar-Salinas K, Seminario JM (2009) Transverse electronic transport in double-stranded DNA nucleotides. *J Phys Chem B* 113:6230–6239. doi:10.1021/jp808790j
69. Sivanesan D, Subramanian V, Nair BU (2001) Solvent effect on DNA base stacked dimers: an isodensity polarizable continuum model approach. *Int J Quantum Chem* 84:750–758
70. Marañón J, Fantoni A, Grigera JR (1999) Adenine-thymine molecular dynamics simulation. Conformation, hydration and magnetic behaviour. *J Mol Liq* 79:177–186
71. Sponer J, Sabat M, Burda JV, Leszczynski J, Hobza P (1999) Interaction of the adenine–thymine Watson–Crick and adenine–adenine reverse-Hoogsteen DNA base pairs with hydrated group IIa (Mg²⁺, Ca²⁺, Sr²⁺, Ba²⁺) and IIb (Zn²⁺, Cd²⁺, Hg²⁺) metal cations: absence of the base pair stabilization by metal-induced polarization effects. *J Phys Chem B* 103:2528–2534. doi:10.1021/jp983744w
72. Coutinho K, Ludwig V, Canuto S (2004) Combined Monte Carlo and quantum mechanics study of the hydration of the guanine–cytosine base pair. *Phys Rev E* 69:061902
73. Gilbert HF (2000) Basic concepts in biochemistry, 2nd edn. McGraw-Hill, Houston
74. Tikhomirova A, Beletskaya IV, Chalikian TV (2006) Stability of DNA duplexes containing GG, CC, AA, and TT mismatches. *Biochemistry* 45:10563–10571. doi:10.1021/bi060304j
75. Hunter WN, Brown T, Anand NN, Kennard O (1986) Structure of an adenine[dot]cytosine base pair in DNA and its implications for mismatch repair. *Nature* 320:552–555

Computer-assisted design for paracetamol masking bitter taste prodrugs

Hatem Hejaz · Rafik Karaman · Mustafa Khamis

Received: 30 January 2011 / Accepted: 14 March 2011 / Published online: 15 April 2011
© Springer-Verlag 2011

Abstract It is believed that the bitter taste of paracetamol, a pain killer drug, is due to its hydroxyl group. Hence, it is expected that blocking the hydroxy group with a suitable linker could inhibit the interaction of paracetamol with its bitter taste receptor/s and hence masking its bitterness. Using DFT theoretical calculations we calculated proton transfers in ten different Kirby's enzyme models, 1–10. The calculation results revealed that the reaction rate is linearly correlated with the distance between the two reactive centers (r_{GM}) and the angle of the hydrogen bonding (α) formed along the reaction pathway. Based on these results three novel tasteless paracetamol prodrugs were designed and the thermodynamic and kinetic parameters for their proton transfers were calculated. Based on the experimental $t_{1/2}$ (the time needed for the conversion of 50% of the reactants to products) and EM (effective molarity) values for processes 1–10 we have calculated the $t_{1/2}$ values for the conversion of the three prodrugs to the parental drug, paracetamol. The calculated $t_{1/2}$ values for ProD 1–3 were found to be 21.3 hours, 4.7 hours and 8 minutes, respectively. Thus, the rate by which the paracetamol prodrug undergoes cleavage to release paracetamol can be determined according to the nature of the linker of the

prodrug (Kirby's enzyme model 1–10). Further, blocking the phenolic hydroxyl group by a linker moiety is believed to hinder the paracetamol bitterness.

Keywords DFT calculations · Kirby's enzyme models · Masking bitter taste · Paracetamol prodrugs · Proton transfer reaction

Introduction

Bitter or unpleasant taste is a major problem in the food and medicine industries. As several oral pharmaceuticals and bulking agents have unpleasant, bitter-tasting components, Pediatric patients resist taking medicine due to their bitterness or unpleasant taste. The bitterness of these preparations leads to lack of patient compliance. The problem of bitterness of drugs in pediatric and geriatric formulations is creating a serious challenge to pharmacists [1].

The desire of improved palatability in bitter taste products has prompted the development of numerous formulations with improved performance and acceptability [2]. In order to satisfy the patient compliance for taking medicines bitterness masking becomes essential. Different approaches are commonly utilized to overcome bitter and unpleasant taste of drugs. This includes reduction of drug solubility in saliva, where a balance between reduced solubility and bioavailability must be achieved, but these approaches were found to be limited and could not overcome the problem of bitterness [3]. Thus, different strategies should be developed in order to overcome this serious problem.

Drugs or molecules interact with taste receptors on the tongue to give bitter, sweet or other taste sensations. Altering the ability of the drug to interact with taste

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1040-5) contains supplementary material, which is available to authorized users.

H. Hejaz · R. Karaman (✉)
Faculty of Pharmacy, Al-Quds University,
P. O. Box 20002, Jerusalem, Palestine
e-mail: dr_karaman@yahoo.com

M. Khamis
Department of Chemistry and Chemical Technology,
Al-Quds University,
P. O. Box 20002, Jerusalem, Palestine

receptors could reduce or eliminate their bitterness. This could be achieved by an appropriate modification of the structure and the size of a bitter compound.

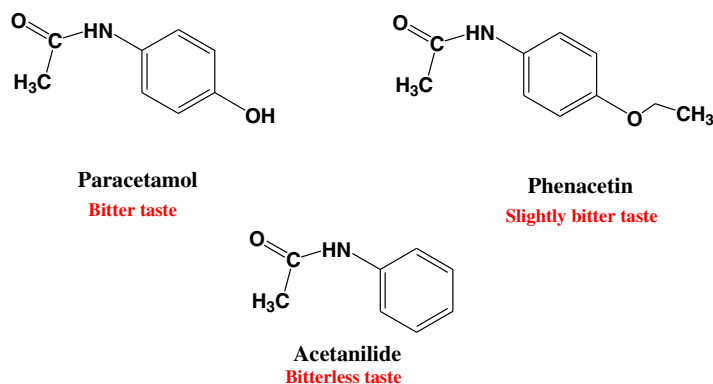
During the last ten years, tremendous progress in the elucidation of bitter taste reception and transduction on the cellular level was made and many new molecules and compounds to reduce off-taste were reported, but none of those were found to eliminate the bitterness fully [4].

Since the discovery and identification of the receptor proteins responsible for bitter taste reception, the mechanism of bitter reception by taste receptor cells seems to be generally known nowadays. Bitter molecules bind to G-protein coupled receptor-type T2R on the apical membrane of the taste receptor cells located in the taste buds. In humans, about 25 different T2R are described. Additionally, several alleles are known and about 1000 different bitter phenotypes exist in human beings [4, 5].

Due to the wide variation of the structure basis of bitter tasting molecules, it is difficult to generalize the molecular requirements for bitterness. Nevertheless, it was noted that a bitter molecule needs a polar group and a hydrophobic moiety (mono polar-hydrophobic concept). A quantitative structure

activity relationship (QSAR) model was developed and could be established for the prediction of bitterness of several analogues. For example, it was reported that an addition of pyridinium moiety to an amino acid chain of a variety of bitter amino acids compounds decreases the bitterness such as in the case of glycine (e.g., pyridinium glycyl betain). Other structural modifications such as an increase in the number of amino groups/residues to more than three and a reduction in the poly-hydroxyl group/COOH has been proven to decrease the bitterness in a significant manner. Furthermore, changing the configuration of bitter compounds by preparing isomer analogues was found to be important for binding affinity to enhance bitterness agonist activity (e.g., L-tryptophan is bitter while D-tryptophan is sweet, and there are many other examples) [4, 6].

Paracetamol is an odorless crystalline compound with a bitter taste widely used as pain killer and to reduce the temperature of patients with fever. These actions are known respectively as analgesic and antipyretic. Initially, paracetamol was found in the urine of patients who had taken phenacetin and later on it was demonstrated that paracetamol was a urinary metabolite of acetanilide [7].



Phenacetin, on the other hand, lacks or has a very slight bitter taste [8]. Examination of the structures of paracetamol and phenacetin reveals that the only difference in the structural features in both is the nature of the group on the *para* position of the benzene ring. While in the case of paracetamol the group is hydroxy, in phenacetin it is ethoxy. Another related example is acetanilide that has a chemical structure similar to that of paracetamol and phenacetin but it lacks the group in the *para* position of the benzene ring. Acetanilide has a burning taste and lacks the bitter taste characteristic for paracetamol [9]. The combined facts described above suggest that the presence of hydroxy group on the *para* position is the major contributor for the bitter taste of paracetamol. Hence, it is

expected that blocking the hydroxy group in paracetamol with a suitable linker could inhibit the interaction of paracetamol with its bitter taste receptor/s and hence masking its bitterness.

It seems reasonable to assume that the phenolic hydroxyl group in paracetamol is crucial for obtaining the bitter taste characteristic for paracetamol. This might be due to the ability of paracetamol to be engaged in a hydrogen bonding net with the active site of its bitter taste receptor *via* its phenolic hydroxyl group.

Recently we have been researching the mechanistic pathways of some intramolecular processes that have been used as enzyme models and prodrug linkers [10–27]. Utilizing DFT and *ab initio* molecular orbital calculation

methods, we have studied: 1) (a) acid-catalyzed lactonization of hydroxy-acids as investigated by Cohen [28–30] and Menger [31–38], (b) S_N2 -based-cyclization reactions of dicarboxylic semi-esters to yield anhydrides as studied by Bruice [39, 40], (c) intramolecular S_N2 -based ring-closing reactions as researched by Brown's group [41] and Mandolini's group [42], (d) proton transfer between two oxygens in Kirby's acetals [43–52], and proton transfer between nitrogen and oxygen in Kirby's enzyme models [43–52], (e) proton transfer between two oxygens in rigid systems as investigated by Menger [31–38], and (f) proton transfer from oxygen to carbon in some of Kirby's enol ethers [43–52] arriving at the following conclusions. (1) The driving force for enhancements in rate for intramolecular processes are both entropy and enthalpy effects. In the cases by which enthalpic effects were predominant such as ring-cyclization and proton transfer reactions proximity or/and steric effects were the driving force for rate accelerations. (2) The nature of the reaction being intermolecular or intramolecular is determined on the distance between the two reactive centers. (3) In S_N2 -based ring-closing reactions leading to three-, four- and five-membered rings the *gem*-dialkyl effect is more dominant in processes involving the formation of an unstrained five-membered ring, and the need for directional flexibility decreases as the size of the ring being formed increases. (4) Accelerations in the rate for intramolecular reactions are a result of both entropy and enthalpy effects. (5) An efficient proton transfer between two oxygens and between nitrogen and oxygen in Kirby's acetal systems were affordable when a strong hydrogen bonding was developed in the products and the corresponding transition states leading to them.

Our previous studies on enzyme models established the necessity to explore the reaction mechanism in order to assign the driving force affecting the reaction rate for a design of an efficient chemical device (prodrug) capable of releasing the parental drug in a programmable manner [10–27]. The prodrugs are designed such that they undergo cleavage reactions in physiological environments (pH 1.5, 6.5 and 7.4) with rates that are completely dependent on the structural features of the inactive linker. A number of linkers could be used to obtain different prodrugs that release the parental drug in different rates that are dependent on the nature or the structural features of the linker.

Continuing our research for utilizing enzyme models as potential linkers for drugs containing hydroxyl groups, we sought to study the mechanism and driving forces affecting the proton transfer rate in some of Kirby's enzyme models (prodrugs linkers). It is expected that such linkers will have a potential to be good carriers to paracetamol. It is worth noting that linking the paracetamol with such linkers *via* its phenolic hydroxyl group will hinder its bitter taste.

Our proposed paracetamol prodrug systems based on proton transfer reactions are illustrated in Scheme 1.

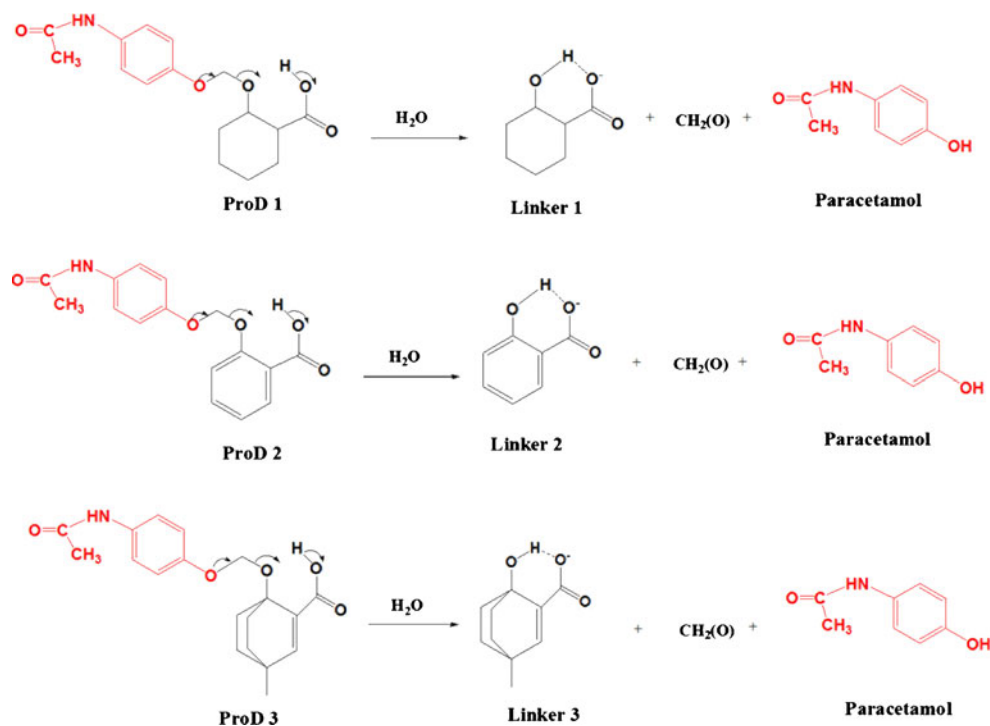
Based on the DFT calculations on the proton transfers in Kirby's enzyme models 1–10 reported herein (Scheme 2), three paracetamol prodrugs are proposed. As shown in Scheme 1, the paracetamol prodrugs, ProD 1–3, have a carboxylic acid group (hydrophilic moiety) and a lipophilic moiety (the rest of the molecule), where the combination of both groups ensures a moderate HLB. It should be noted that the HLB value will be determined upon the physiologic environment by which the prodrug is dissolved. For example, in the stomach, the paracetamol prodrugs will primarily exist in the carboxylic acid form whereas in the blood circulation the carboxylate anion form will be predominant. It is planned that ProD 1–3 will be obtained as sodium or potassium carboxylate salt since this form is expected to be resistant to cleavage by a proton transfer process. The latter is feasible when the carboxylate anion is converted to the free acid form (Scheme 1).

It should be emphasized that prodrugs ProD 1–3 could be used in different dosage forms (tablets, syrups, suppositories, and etc.) because of their potential solubility in organic and aqueous media due to the ability of their carboxyl groups to be converted to the corresponding carboxylate salt.

In this manuscript, we describe our DFT quantum molecular orbital investigations of ground state and transition state structures, vibrational frequencies, and reaction trajectories for intramolecular proton transfer in ten of Kirby's enzyme model systems 1–10 (Scheme 2). It is expected that the calculations study on systems 1–10 will provide a good basis for the prediction of the pharmacokinetic behavior of paracetamol prodrugs of the type ProD 1–3.

Calculations methods

The DFT calculations at B3LYP/6-31G (d,p) level were carried out using the quantum chemical package Gaussian-98 [53]. The starting geometries of all the molecules presented in this study were obtained using the Argus Lab program [54] and were initially optimized in the presence of one molecule of water at the AM1 level of theory, followed by an optimization at the HF/6-31G level [53]. The calculations were carried out based on the restricted Hartree-Fock (RHF) method with full optimization of all geometrical variables. An energy minimum (a stable compound or a reactive intermediate) has no negative vibrational force constant. A transition state is a saddle point which has only one negative vibrational force constant [55]. The "reaction coordinate method" [56] was used to calculate the activation energy in systems 1–10 and ProD 1–3. In this method, one bond length is constrained for the appropriate degree of freedom while all other variables are freely optimized. The activation energy values for the

Scheme 1 Proton transfer reactions in ProD 1–3

proton transfer reactions were calculated from the difference in energies of the global minimum structures and the derived transition states (TS). Verification of the desired reactants and products was accomplished using the “intrinsic coordinate method” [56]. The transition state structures were verified by their only one negative frequency. Full optimization of the transition states was accomplished after removing any constraints imposed while executing the energy profile. The activation energies obtained from DFT at B3LYP/6-31G (d,p) level of theory for 1–10 and ProD 1–3 were calculated with and without the inclusion of solvent (water). The calculations with the incorporation of a solvent were performed using the integral equation formalism model of the polarizable continuum model (PCM) [57–60].

Results and discussion

To mask paracetamol bitterness, the pro-drug approach of linking paracetamol to a moiety, which upon dissolution in a physiologic environment, liberates the parental drug seems to be promising. Kirby’s mechanistic study on acetals that were used as enzyme models has inspired us to exploit these models as appropriate linkers for masking the bitter taste of paracetamol. Kirby’s kinetic study on processes 1–10 (Scheme 2) indicates that the rate-limiting step in these processes is a proton transfer from the carboxylic hydroxyl group into the neighboring ether oxygen. In addition, Kirby’s findings revealed that the proton transfer rate is largely determined on the strength of the hydrogen

bonds in the products and consequently in the transition state leading to them [43–52]. Thus, it is safe to conclude that the proton transfer rate is dependent on the structural features of the enzyme model (linker) as evident from the different experimental rate values measured for processes 1–10.

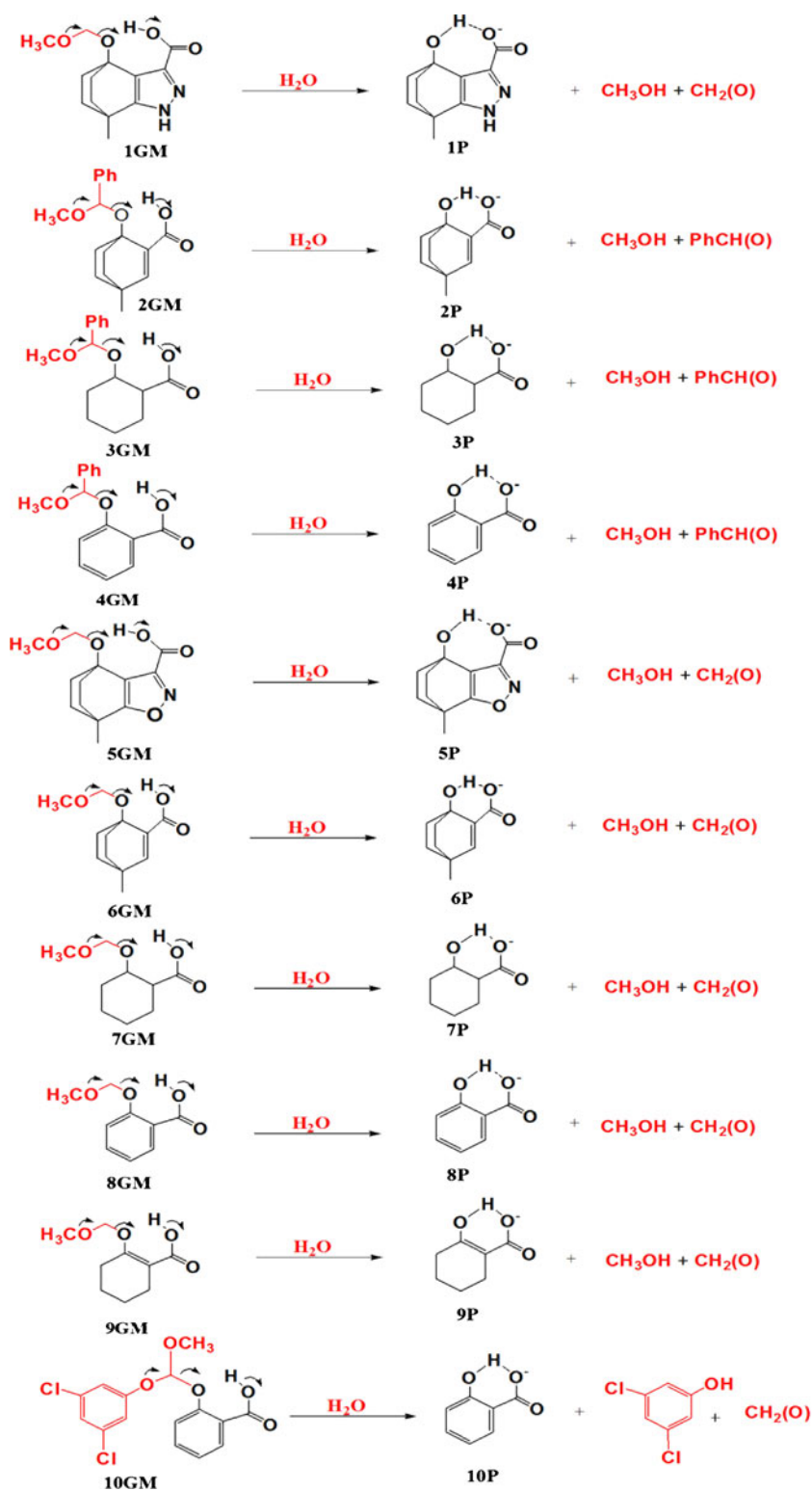
Replacing the leaving group (methoxy) in the reactants of the linkers 1–10 (Scheme 2) with paracetamol (ProD 1–3, Scheme 1) is not expected to have any significant effect on the rates of the proton transfer in these processes. Therefore, computational investigation of the kinetic properties for these linkers will shed light on the cleavage rates of paracetamol prodrugs ProD 1–3.

General consideration

Because the energy of carboxylic acid is strongly dependent on its conformation and especially on its ability to be engaged either inter- or intramolecularly in hydrogen bonding, we were concerned with the identification of the most stable conformation (global minimum) for each of Kirby’s enzyme models 1–10 and prodrugs ProD 1–3 calculated in this study. This was accomplished by 36 rotations of the carboxyl group about the bond C4–C6 in increments of 10° (i.e., variation of the dihedral angle O5C4C6C7, see Chart 1) and calculation of the energies of the resulting conformers.

In the DFT calculations of the reactants in 1–10 and ProD 1–3, two types of conformers were considered: one in which the carboxylic hydroxyl proton is *syn* to the alkoxy

Scheme 2 Proton transfer reactions in 1–10, where GM and P are the reactants and products, respectively



group in the β position of the carboxylic acid moiety (R, see Chart 1) and another in which it is *anti*. It was found that all global minimum structures exhibit a *syn* conformation. Further, the calculation results revealed that in 3, 7, 10 and ProD 1 the global minimum structures reside in a

conformation by which the carboxylic hydroxyl proton is far away from the alkoxy oxygen (no hydrogen bonding exists between the carboxylic hydroxyl proton and the alkoxy oxygen) and instead it engages in a hydrogen bonding with a molecule of water (Fig. 1a). For systems 1–2,

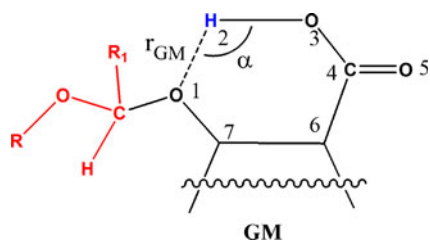


Chart 1 Schematic representation of the reactants in the proton transfers of Kirby's enzyme models 1–10 and ProD 1–3. GM is the global minimum structure, r_{GM} is the O–H distance in the GM. α , is the angle of attack (hydrogen bonding) O–H–O in the GM

4–6, 8–9, and ProD 2–3, the global minimum conformations were those having intramolecular hydrogen bonding between the carboxylic hydroxyl proton and the alkoxy oxygen (see Fig. 1a).

Structural analysis for the entities involved in the proton transfers of 1–10 and ProD 1–3

Global minimum structures (GM)

Generally, reactions in aqueous medium involve interactions between the reactants and water. The nature for such interactions is dependent on the structural features of the reactants. Since the proton transfers in 1–10 were carried out in aqueous medium, we have calculated the geometries of the entities involved in these processes in the presence of one molecule of water followed by optimization in a cluster of water. Figure 1a and Table 1 illustrate the DFT calculated properties for the global minimum structures of 1–10 (1GM–10GM) and ProD 1–3 (ProD 1GM–3GM). Examination of the optimized structures for 1GM–10GM and ProD 1GM–3GM indicate that 3GM, 7GM, 10GM, and ProD 1GM exhibit conformations by which the carboxylic group is engaged intermolecularly in a hydrogen bonding with a molecule of water. This is because the carboxyl group in 3GM, 7GM, 10GM, and ProD 1GM prefers to be engaged in hydrogen bonding with a molecule of water rather than intramolecularly, since the latter is energetically expensive due to a high energy barrier for the rotation of the carboxyl group around the C4–C6 bond [43–52]. It should be indicated that Fife reported that acetal 3GM shows no intramolecular general acid catalysis by the neighboring carboxyl group [61].

On the other hand, 1GM–2GM, 4GM–6GM, 8GM–9GM and ProD 2GM–3GM were found to reside in a conformation by which the carboxylic hydroxyl proton is engaged in an intramolecular hydrogen bond with the neighboring alkoxy oxygen. This engagement results in the formation of a seven-membered ring for 1GM and 5GM and a six-membered ring for 2GM–4GM, 6GM, 8GM, 9GM, ProD

2GM, and ProD 3GM (see Figs. 1a and 2). The DFT calculated hydrogen bonding length (r_{GM}) in 3GM, 7GM, 10GM, and ProD 1GM (reactants that are engaged intermolecularly with a water molecule) was found to be in the range of 3.60 Å–3.66 Å and that for the attack angle α (the hydrogen bond angle, O1H2O3) to be in the range of 47°–58°. On the other hand, the r_{GM} and α values for 1GM–10GM and ProD 1GM–3GM were 1.69 Å–1.75 Å and 143°–170.7°, respectively. Furthermore, the hydrogen bonding strength, r_{GM} (O1–H2), varies according to the structural features of the starting geometry.

Transition state structures (TS)

The DFT calculated geometries for the transition states of 1–10 (1TS–10TS) and ProD 1–3 (ProD 1TS–3TS) are depicted in Figs. 1b and 2 and Table 1. Inspection of the optimized structures for 1TS–10TS and ProD 1TS–3TS revealed that all the TS structures involve strong hydrogen bonding between the carboxylic hydroxyl proton (H2) and the ether oxygen. The calculated hydrogen bonding angle (O1H2O3) in the TS optimized structures was found in the range of 131°–170°.

Product geometries (P)

The DFT calculated geometries for the products of 1–10 (1P–6P) and ProD 1–3 (ProD 1P–3P) shown in Fig. 1c indicate the presence of a strong hydrogen bond between O1H2O3 where the bond length was found in the range 1.45 Å to 1.66 Å and the O1H2O3 angle was in the range of 154°–170°. This is similar to that found for the corresponding transition state structures, 1TS–10TS and ProD 1TS–3TS.

Calculations of the kinetic properties for the proton transfer in 1–10 and ProD 1–3

Using the quantum chemical package Gaussian-98 [53] we calculated the DFT at B3LYP/6-31G (d,p) level of theory the kinetic and thermodynamic properties for the intramolecular proton transfer in processes 1–10 and ProD 1–3 (Schemes 1 and 2).

The B3LYP/6-31G (d,p) activation energies were calculated in the presence of one molecule of water and with the inclusion of a cluster of water as a solvent. The calculation results show that the energies with and without a cluster of water are significantly different (Table 2). This indicates that the presence of water as a solvent has a profound effect on the proton transfer rate values. This is in accordance with previously reported studies by Kirby and Fife that show the importance of water in the mechanistic pathway for the proton transfer in such systems [43–52, 61].

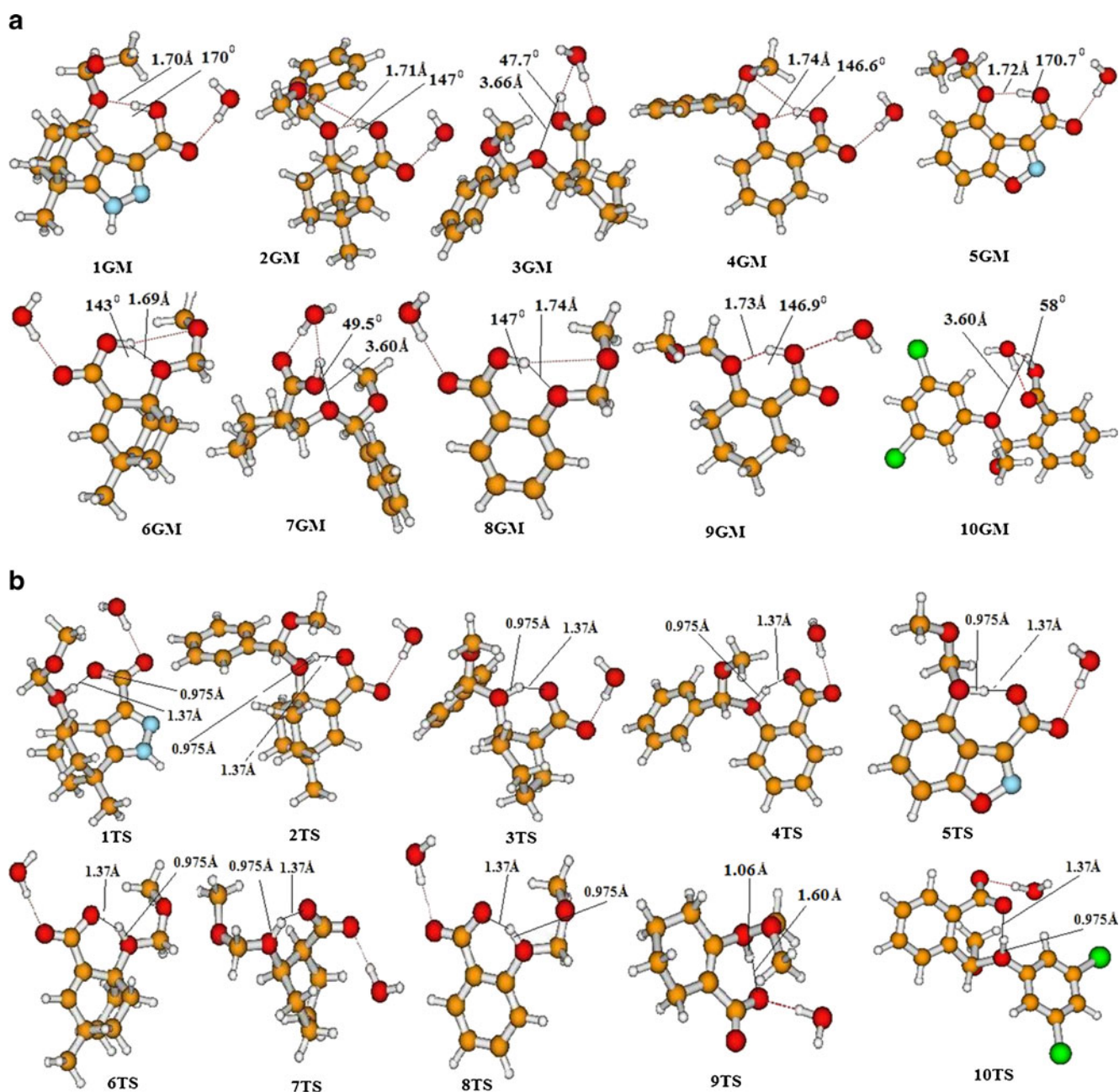


Fig. 1 (a) DFT optimized structures for the global minimum (GM) structures in the intramolecular proton transfer reaction of 1–10. (b) DFT optimized structures for the transition state (TS) structures in the

intramolecular proton transfer reaction of 1–10. (c) DFT optimized structures for the product (P) structures in the intramolecular proton transfer reaction of 1–10

Using the calculated DFT enthalpic and entropic energies for the global minimum structures 1GM–10GM and ProD 1GM–3GM and their corresponding transition states 1TS–10TS and ProD 1TS–3TS (Table 1) we have calculated the enthalpic (ΔH^\ddagger), the entropic ($T\Delta S^\ddagger$), and the free activation energies (ΔG^\ddagger), in the gas phase and in the presence of cluster of water, for the corresponding proton transfer reactions. The calculated energy values are summarized in Table 2.

The effect of the distance r_{GM} and the angle α on the activation energy for the proton transfer in processes 1–6

The intramolecular hydrogen bonding (r_{GM}) between the hydroxyl proton (H2) and the ether oxygen O1 (O1–H2) was found to be dependent on the conformation of the global minimum structure. Short r_{GM} values were obtained when the attack angle (α) values in the GM conformations

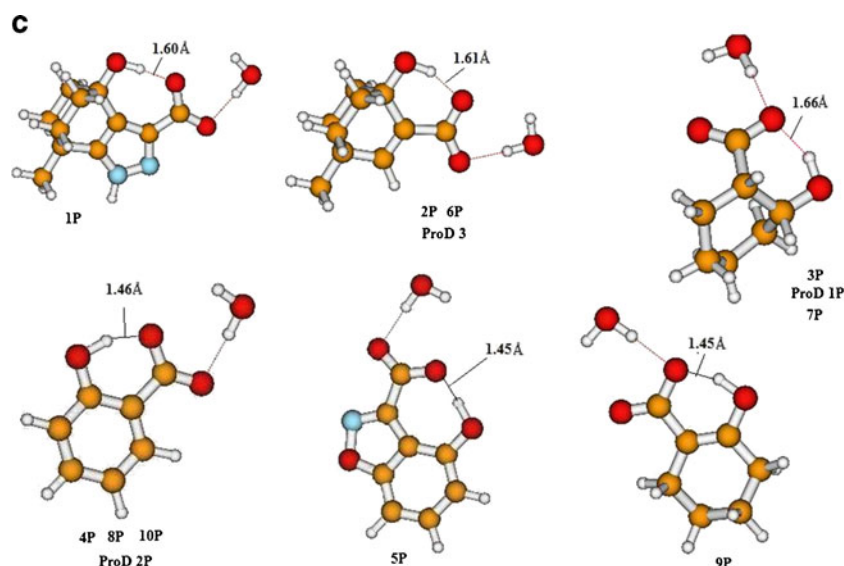


Fig. 1 (continued)

were high and close to 180° , whereas small α values resulted in longer r_{GM} distances (Figs. 1a and 2). In fact, a strong correlation with a correlation coefficient of $R=0.98$ was found between the distance r_{GM} and the attack angle α (Fig. 3a).

Examination of Table 2 revealed that the calculated free activation energy (ΔG^\ddagger) in the gas phase and in water needed to execute a proton transfer in systems 1–10 and ProD 1–3 is largely affected by both the O1-H2 distance (r_{GM}) and the attack angle O1H2O3 (α) (see Chart 1). Acetals having global minimum conformation with low r_{GM} and high α , such as 1, 5, and ProD 3, undergo proton

transfer in higher rates (lower ΔG^\ddagger) than those having high r_{GM} and low α values, such as 3, 7, and ProD 1. When r_{GM} and α values were examined for correlation with the calculated DFT enthalpic energies (ΔH^\ddagger) a linear correlation was found between ΔH^\ddagger and $r_{GM}^2 \times \sin(180-\alpha)$ with a correlation coefficient of $R=0.95$. On the other hand, a correlation of the activation free energies (ΔG^\ddagger) with $r_{GM}^2 \times \sin(180-\alpha)$ gave an R value of 0.85 (Fig. 3b).

The measure generally used for intramolecular efficiency is the effective molarity (EM). The EM parameter is defined as k_{intra}/k_{inter} for corresponding intramolecular and intermolecular processes driven by identical mechanisms.

Table 1 DFT (B3LYP) calculated properties for the proton transfer reactions of 1–10 and ProD 1–3

Struct.	Enthalpy, H In Hartree	Entropy, S, Cal/Mol-Kelvin	Frequency Cm^{-1}	Struct.	Enthalpy, H In Hartree	Entropy, S, Cal/Mol-Kelvin	Frequency Cm^{-1}
1GM	-993.05253	153.17	—————	8GM	-726.31205	132.92	—————
1TS	-993.00826	144.17	181.03i	8TS	-726.26559	119.33	132.40i
2GM	-1076.50693	172.52	—————	9GM	-728.72138	133.20	—————
2TS	-1076.45691	160.07	611.61i	9TS	-728.66450	120.34	522.39i
3GM	-957.37381	159.64	—————	10GM	-1876.55809	152.33	—————
3TS	-957.32753	142.46	184.58i	10TS	-1876.49684	154.58	719.08i
4GM	-961.00056	152.14	—————	ProD 1GM	-1129.71308	171.26	—————
4TS	-960.93507	155.62	578.23i	ProD 1TS	-1129.63371	160.51	612.59i
5GM	-893.73201	135.57	—————	ProD 2GM	-1126.07858	165.90	—————
5TS	-893.68532	135.66	165.49i	ProD 2TS	-1126.02298	160.67	215.70i
6GM	-845.44633	147.15	—————	ProD 3GM	-1245.20203	173.32	—————
6TS	-845.39359	142.97	558.92i	ProD 3TS	-1245.15174	175.96	256.26i
7GM	-729.94298	129.27	—————	—————	—————	—————	—————
7TS	-729.87241	121.80	390.28i	—————	—————	—————	—————

B3LYP refers to values calculated by B3LYP/6-31G (d, p) in the gas phase. GM and TS are global minimum and transition state structures, respectively

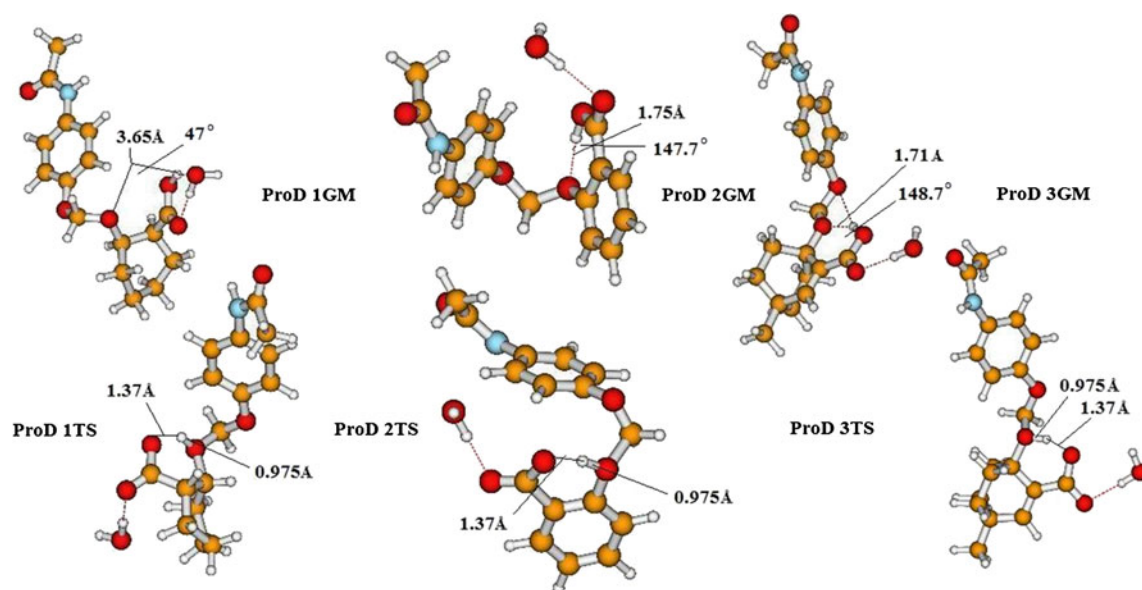


Fig. 2 DFT optimized global minimum (GM) and transition state structures (TS) for processes ProD 1–3

The factors determining the EM value are ring size, solvent and reaction type. Cyclization reactions *via* intramolecular nucleophilic addition are much more efficient than intramolecular proton transfers. Values in the order of 10^9 – 10^{13} M have been measured for the effective molarity in intramolecular processes occurring through nucleophilic addition. Whereas for proton transfer processes values of less than 10 M were obtained [62]. The effective molarity parameter is considered an excellent tool to describe the efficiency of a certain intramolecular process.

In order to obtain credibility for our calculation results we introduce our computation rational for calculating the

EM values for processes 1–10 based on the DFT calculated activation energies (ΔG^\ddagger) of 1–10 and the corresponding intermolecular process 11 (Scheme 3).

Equation 5 which describes the EM term as a function of the difference in the activation energies of the intra- and the corresponding intermolecular processes was derived from Eqs. 1, 2, 3 and 4. The calculated EM values for processes 1–10 by using Eq. 5 are listed in Table 2.

$$EM = k_{\text{intra}}/k_{\text{inter}} \quad (1)$$

$$\Delta G_{\text{inter}}^\ddagger = -RT \ln k_{\text{inter}} \quad (2)$$

Table 2 DFT (B3LYP/6-31G (d,p)) calculated kinetic and thermodynamic properties for the proton transfers in 1–10 and ProD 1–3

System	log EM ^{43–52} (exp.)	log EM (calc.)	ΔH^\ddagger (gas phase)	T ΔS^\ddagger (gas phase)	ΔG^\ddagger (gas phase)	ΔH^\ddagger (Water)	ΔG^\ddagger (Water)
1	10.000	10.58	27.78	-2.68	30.46	21.47	24.15
2	3.4771	6.04	31.38	-3.71	35.09	26.64	30.35
3	—————	-0.41	41.09	1.04	40.05	40.15	39.11
4	3.9777	5.30	29.04	-5.12	34.16	26.22	31.34
5	12.600	12.72	29.30	0.03	29.27	21.22	21.19
6	—————	7.40	33.09	-1.25	34.34	24.71	25.96
7	—————	-1.52	44.29	-2.23	46.52	34.76	36.99
8	4.0000	5.17	34.32	-3.42	37.74	27.53	31.52
9	4.3010	5.30	35.69	-3.83	39.52	27.49	31.32
10	1.5798	-1.52	38.43	0.67	37.76	37.66	36.99
ProD 1	—————	-5.81	49.80	-3.20	53.00	39.65	42.85
ProD 2	—————	6.55	34.89	-1.56	36.45	28.09	29.65
ProD 3	—————	11.75	31.56	0.79	29.77	23.30	22.51

ΔH^\ddagger is the activation enthalpic energy (kcal/mol). T ΔS^\ddagger is the activation entropic energy in kcal/mol. ΔG^\ddagger is the activation free energy (kcal/mol). log EM (exp.) and log EM (calc.) are the experimental and calculated effective molarities ($EM = k_{\text{intra}}/k_{\text{inter}}$)

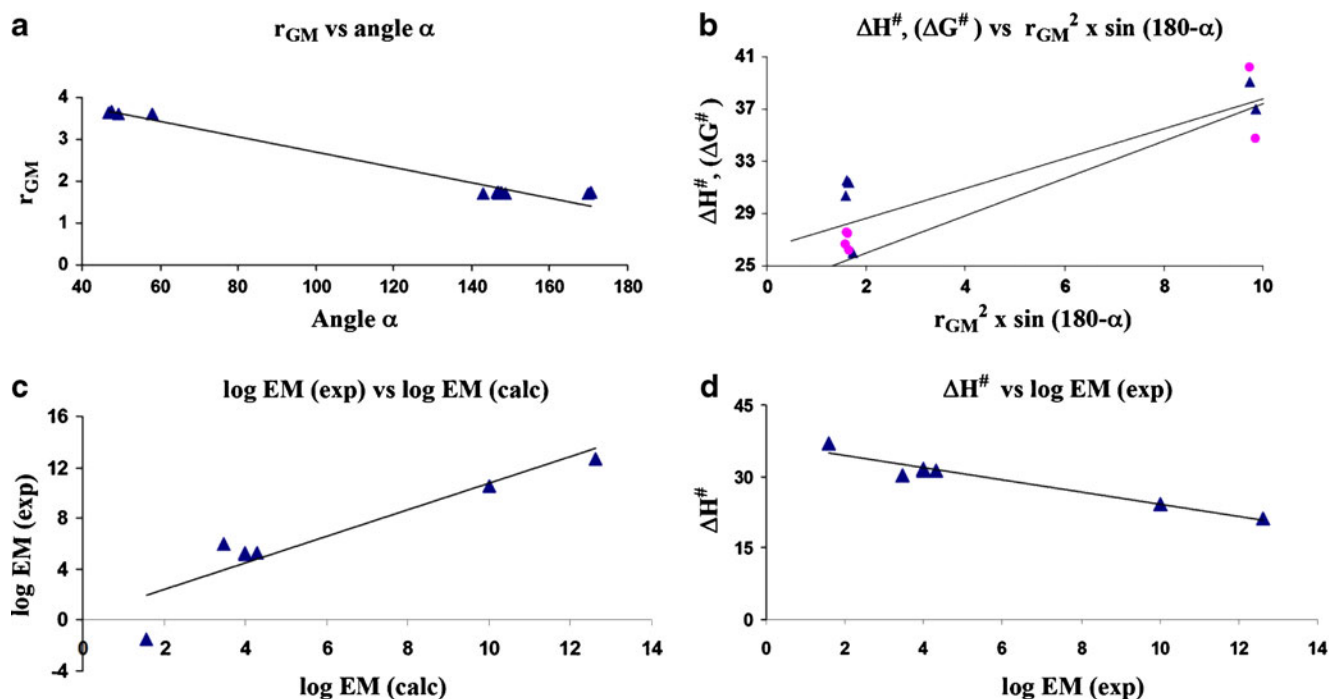


Fig. 3 (a) Plot of the DFT calculated r_{GM} vs. angle α in 1–10 and ProD 1–3, where r_{GM} and α are the distance between the two reactive centers and the attack (hydrogen bond) angle in the GM structure, respectively. (b) Plot of the DFT calculated ΔH^\ddagger and ΔG^\ddagger vs. $r_{GM}^2 \times \sin(180-\alpha)$

ΔH^\ddagger vs. $r_{GM}^2 \times \sin(180-\alpha)$ in 1–10 (c) Plot of the experimental EM values vs. the calculated EM values in 1–10 (d) Plot of the DFT calculated ΔH^\ddagger values vs. the experimental EM values in 1–10

$$\Delta G_{intra}^\ddagger = -RT \ln k_{intra} \quad (3)$$

$$\Delta G_{intra}^\ddagger - \Delta G_{inter}^\ddagger = -RT (\ln k_{intra}/k_{inter}) \quad (4)$$

$$EM = e^{-(\Delta G_{inter}^\ddagger - \Delta G_{intra}^\ddagger)/RT}, \quad (5)$$

where T is 298° K and R is the gas constant.

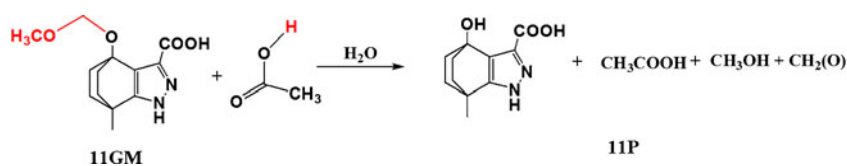
Careful inspection of the effective molarity values shown in Table 2 revealed that systems 1, 5, and ProD 3 are the most efficient processes among 1–10 and ProD 1–3 ($\log EM > 10$) and the least efficient are 3, 7, 10, and ProD 1 with $\log EM < 1.6$. Furthermore, the calculated EM values were examined for correlation with the experimental EM values. The correlation results show good correlation between the two parameters with a correlation coefficient of $R=0.92$ (Fig. 3c). Similarly, good correlation results with a correlation coefficient of $R=0.97$ were obtained when the

enthalpic activation energy values were correlated with the experimental EM values (Fig. 3d) [16, 23].

Examination of the calculated EM values for ProD 1–3 (Table 2) indicates that ProD 3 is the process with the highest rate whereas ProD 1 is the process with the lowest rate among the three prodrugs. Based on the experimental $t_{1/2}$ (the time needed for the conversion of 50% of the reactants to products) values for processes 1 and 5 and the EM values for 1, 5, and ProD 1–3 we have calculated the $t_{1/2}$ values for the conversion of the three prodrugs to the parental drug, paracetamol. The calculated $t_{1/2}$ values for ProD 1–3 are 21.3 hours, 4.7 hours and 8 minutes, respectively.

Based on these results solid and liquid preparations of three prodrugs could be made in order to achieve two goals: (1) masking the bitterness of paracetamol and (2) slow release of paracetamol where the rate of the release is determined on the nature of the prodrug linker (Kirby's enzyme model).

Scheme 3 Intermolecular proton transfer in 11, where GM and P are the reactants and products, respectively



Summary and conclusions

The DFT calculation results revealed that the activation energy for the proton transfer in processes 1–10 and ProD 1–3 is quite dependent on the geometric variations in the ground state. It was found that the distance between the two reactive centers, r_{GM} , and the angle of attack α are the main factors affecting the rate (activation energy) of the proton transfer. A short distance between the two reactive centers (r_{GM}) and hydrogen bonding angle close to 180° result in strong intramolecular hydrogen bonding that clearly indicates the directionality of hydrogen bonding, hence to stabilization of the product and the transition states leading to them and consequently to accelerations in rate.

The linear correlation of the experimental and calculated EM values established a good basis for the design of some paracetamol prodrugs that are capable of masking the bitter taste of paracetamol and have the potential to be used in slow release preparations. Further, this study could be utilized to design other prodrugs by which the rate for the release of the parental drug could be determined according to the nature of the linker (Kirby's enzyme model).

Our future directions include the synthesis of paracetamol prodrugs ProD 1–3 according to known procedures. *in vitro* kinetic studies at different pH values will be made in order to be utilized for the *in vivo* pharmacokinetic studies.

Based on the *in vitro* results, the prodrugs will be tested *in vivo* in addition to paracetamol as a control. The prodrug will be administered to animals by I.V. injection and per-os, blood and urine samples will be collected at different times. The concentration of paracetamol will be determined using a reliable bioanalytical method. Further, pharmacokinetic parameter values will be calculated including oral bioavailability, terminal elimination half-life and other pharmacokinetic parameters as deemed necessary.

Acknowledgments The Karaman Co. is thanked for support of our computational facilities. Special thanks are also given to Angi Karaman, Donia Karaman, Rowan Karaman and Nardene Karaman for technical assistance.

References

- Remington RWJ (2002) The science and practice of pharmacy, 20th edn. Mack publishing company, Easton, pp 1018–1020
- Brahmankar DM, Jaiswal SB (1995) Biopharmaceutics & pharmacology, 1st edn. Vallabh Prakashan, Delhi, pp 162–165
- Kuchekar BS, Badhan AC, Mahajan HS (2003) Mouth dissolving tablets: a novel drug delivery system. *Pharma Times* 35:7–9
- Ley JP (2008) Masking bitter taste by molecules. *Chem Precept Chem Precept* 1:58–77
- Chandreshekar J, Mueller K, Hoon MA, Adler E, Feng L, Guo W, Zuker CS, Ryba NJP (2000) T2Rs function as bitter taste receptor. *Cell (Cambridge, Mass)* 100:703–711
- Scotti L, Scotti MT, Ishiki HM, Ferreira MGP, Emerenciano VP, Menezes CMS, Ferreira EI (2007) Quantitative elucidation of the structure-bitterness relationship of cynaropicrin and grosheimin derivatives. *Food Chem* 105:77–83
- <http://www.assistpainrelief.com/dyn/304/Paracetamol.html>
- http://www.chemicalbook.com/ChemicalProductProperty_EN_CB6141828.htm
- <http://www.chemicalall.com/chemicals-name-a/acetanilide.html>
- Karaman R (2008) Analysis of Menger's spatiotemporal hypothesis. *Tetrahedron Lett* 49:5998–6002
- Karaman R (2009) A new mathematical equation relating activation energy to bond angle and distance: a key for understanding the role of acceleration in the lactonization of the trimethyl lock system. *Bioorg Chem* 37:11–25
- Karaman R (2009) Reevaluation of Bruice's proximity orientation. *Tetrahedron Lett* 50:452–456
- Karaman R (2009) Accelerations in the lactonization of trimethyl lock systems is due to proximity orientation and not to strain effects. *Res Lett Org Chem*. doi:10.1155/2009/240253, 5 pages
- Karaman R (2009) The effective molarity (EM) puzzle in proton transfer reactions. *Bioorg Chem* 37:106–110
- Karaman R (2009) Cleavage of Menger's aliphatic amide: a model for peptidase enzyme solely explained by proximity orientation in intramolecular proton transfer. *J Mol Struct THEOCHEM* 910:27–33
- Karaman R (2009) The *gem*-disubstituent effect-computational study that exposes the relevance of existing theoretical models. *Tetrahedron Lett* 50:6083–6087
- Karaman R (2010) Effects of substitution on the effective molarity (EM) for five membered ring-closure reactions- a computational approach. *J Mol Struct THEOCHEM* 939:69–74
- Karaman R (2009) Analyzing Kirby's amine olefin – a model for amino-acid ammonia lyases. *Tetrahedron Lett* 50:7304–7309
- Karaman R (2010) The effective molarity (EM) puzzle in intramolecular ring-closing reactions. *J Mol Struct THEOCHEM* 940:70–75
- Karaman R (2010) The efficiency of proton transfer in Kirby's enzyme model, a computational approach. *Tetrahedron Lett* 51:2130–2135
- Karaman R (2010) Proximity vs. strain in ring-closing reactions of bifunctional chain molecules- a computational approach. *J Mol Phys* 108:1723–1730
- Karaman R (2010) The effective molarity (EM) – a computational approach. *Bioorg Chem* 38:165–172
- Karaman R (2010) A general equation correlating intramolecular rates with “attack” parameters distance and angle. *Tetrahedron Lett* 51:5185–5190
- Karaman R, Alfalah S (2010) Multi transition states in SN2 intramolecular reactions. *Int Rev Biophys Chem* 1:14–23
- Karaman R, Pascal R (2010) A computational analysis of intramolecularity in proton transfer reactions. *Org Bimol Chem* 8:5174–5178
- Karaman R, Hallak H (2010) Anti-malarial pro-drugs- a computational aided design. *Chem Biol Drug Des* 76:350–360
- Karaman R (2010) Prodrugs of Aza nucleosides based on proton transfer reactions. *J Comput Mol Des* 24:961–970
- Milstein S, Cohen LA (1970) Concurrent general-acid and general-base catalysis of esterification. *J Am Chem Soc* 92:4377–4382
- Milstein S, Cohen LA (1970) Rate acceleration by stereopopulation control: models for enzyme action. *Proc Natl Acad Sci USA* 67:1143–1147
- Milstein S, Cohen LA (1972) Stereopopulation control. I. Rate enhancement in the lactonizations of *o*-hydroxyhydrocinnamic acids. *J Am Chem Soc* 94:9158–9165
- Menger FM, Ladika M (1990) Remote enzyme-coupled amine release. *J Org Chem* 35:3006–3007

32. Menger FM, Ladika M (1988) Fast hydrolysis of an aliphatic amide at neutral pH and ambient temperature. A peptidase model. *J Am Chem Soc* 110:6794–6796
33. Menger FM (1985) On the source of intramolecular and enzymatic reactivity. *Acc Chem Res* 18:128–134
34. Menger FM, Chow JF, Kaiserman H, Vasquez PC (1983) Directionality of proton transfer in solution. Three systems of known angularity. *J Am Chem Soc* 105:4996–5002
35. Menger FM (1983) Directionality of organic reactions in solution. *Tetrahedron* 39:1013–1040
36. Menger FM, Grossman J, Liotta DC (1983) Transition-state pliability in nitrogen-to-nitrogen proton transfer. *J Org Chem* 48:905–907
37. Menger FM, Galloway AL, Musaev DG (2003) Relationship between rate and distance. *Chem Commun* 2370–2371
38. Menger FM (2005) An alternative view of enzyme catalysis. *Pure Appl Chem* 77:1873–1876, and references therein
39. Bruice TC, Pandit UK (1960) The effect of geminal substitution ring size and rotamer distribution on the intramolecular nucleophilic catalysis of the hydrolysis of monophenyl esters of dibasic acids and the solvolysis of the intermediate anhydrides. *J Am Chem Soc* 82:5858–5865
40. Bruice TC, Pandit UK (1960) Intramolecular models depicting the kinetic importance of “Fit” in enzymatic catalysis. *Proc Natl Acad Sci USA* 46:402–404
41. Brown RF, van Gulick NM (1956) The geminal alkyl effect on the rates of ring closure of bromobutylamines. *J Org Chem* 21:1046–1049
42. Galli C, Mandolini L (2000) The role of ring strain on the ease of ring closure of bifunctional chain molecules. *Eur J Org Chem* 3117–3125 and references therein
43. Kirby AJ, Parkinson A (1994) Most efficient intramolecular general acid catalysis of acetal hydrolysis by the carboxyl group. *J Chem Soc Chem Commun* 707–708
44. Brown CJ, Kirby AJ (1997) Efficiency of proton transfer catalysis. Intramolecular general acid catalysis of the hydrolysis of dialkyl acetals of benzaldehyde. *J Chem Soc Perkin Trans* 2:1081–1093
45. Craze G-A, Kirby AJ (1974) The hydrolysis of substituted 2-methoxymethoxybenzoic acids. *J Chem Soc Perkin Trans* 2:61–66
46. Barber SE, Dean KES, Kirby AJ (1999) A mechanism for efficient proton-transfer catalysis. Intramolecular general acid catalysis of the hydrolysis of 1-arylethyl ethers of salicylic acid. *Can J Chem* 79:792–801
47. Kirby AJ, de Silva MF, Lima D, Roussev CD, Nome F (2006) Efficient intramolecular general acid catalysis of nucleophilic attack on a phosphodiester. *J Am Chem Soc* 128:16944–16952
48. Kirby AJ, Williams NH (1994) Efficient intramolecular general acid catalysis of enol ether hydrolysis. Hydrogen-bonding stabilization of the transition state for proton transfer to carbon. *J Chem Soc Perkin Trans* 2:643–648
49. Kirby AJ, Williams NH (1991) Efficient intramolecular general acid catalysis of vinyl ether hydrolysis by the neighbouring carboxylic acid group. *J Chem Soc Chem Commun* 1643–1644
50. Hartwell E, Hodgson DRW, Kirby AJ (2000) Exploring the limits of efficiency of proton-transfer catalysis in models and enzymes. *J Am Chem Soc* 122:9326–9327
51. Kirby AJ (1997) Efficiency of proton transfer catalysis in models and enzymes. *Acc Chem Res* 30:290–296
52. Asaad N, Davies JE, Hodgson DRW, Kirby AJ (2005) The search for efficient intramolecular proton transfer from carbon: the kinetically silent intramolecular general base-catalysed elimination reaction of o-phenyl 8-dimethylamino-1-naphthaldoximes. *J Phys Org Chem* 18:101–109
53. <http://www.gaussian.com>
54. Casewit CJ, Colwell KS, Rappe AK (1992) Application of a universal force field to main group compounds. *J Am Chem Soc* 114:10046–10053
55. Murrell JN, Laidler KJ (1968) Symmetries of activated complexes. *Trans Faraday Soc* 64:371–377
56. Muller K (1980) Reaction paths on multidimensional energy hypersurfaces. *Angew Chem Int Ed Engl* 19:1–13
57. Cancès MT, Mennucci B, Tomasi J (1997) A new integral equation formalism for the polarizable continuum model: theoretical background and applications to isotropic and anisotropic dielectrics. *J Chem Phys* 107:3032–3041
58. Mennucci B, Tomasi J (1997) Continuum solvation models: a new approach to the problem of solute’s charge distribution and cavity boundaries. *J Chem Phys* 106:5151–5158
59. Mennucci B, Cancès MT, Tomasi J (1997) Evaluation of solvent effects in isotropic and anisotropic dielectrics and in ionic solutions with a unified integral equation method: theoretical bases, computational implementation, and numerical applications. *J Phys Chem B* 101:10506–10517
60. Tomasi J, Mennucci B, Cancès MT (1997) The IEF version of the PCM solvation method: an overview of a new method addressed to study molecular solutes at the QM *ab initio* level. *J Mol Struct THEOCHEM* 464:211–226
61. Fife TH, Przystas TJ (1979) Intramolecular general acid catalysis in the hydrolysis of acetals with aliphatic alcohol leaving groups. *J Am Chem Soc* 101:1202–1210
62. Kirby AJ (2005) Effective molarities for intramolecular reactions. *J Phys Org Chem* 18:101–278

Comparative modeling of UDP-N-acetylmuramoyl-glycyl-D-glutamate-2, 6-diaminopimelate ligase from *Mycobacterium leprae* and analysis of its binding features through molecular docking studies

Anusuya Shanmugam · Jeyakumar Natarajan

Received: 28 December 2010 / Accepted: 14 March 2011 / Published online: 15 April 2011
© Springer-Verlag 2011

Abstract Leprosy is an infectious disease caused by *Mycobacterium leprae*. The increasing drug and multi-drug resistance of *M. leprae* enforce the importance of finding new drug targets. Mycobacterium has unusually impermeable cell wall that contributes to considerable resistance to many drugs. Peptidoglycan is an important component of the cell wall of *M. leprae*. UDP-N-acetylmuramoyl-glycyl-D-glutamate-2, 6-diaminopimelate ligase (MurE) plays a crucial role in the peptidoglycan biosynthesis and hence it could be considered as a potential drug target for leprosy. Structure of this enzyme for *M. leprae* has not yet been elucidated. We modeled the three-dimensional structure of MurE from *M. leprae* using comparative modeling methods based on the X-ray crystal structure of MurE from *E. coli* and validated. The 3D-structure of *M. leprae* MurE enzyme was docked with its substrates *meso*-diaminopimelic acid (A₂pm) and UDP-N-acetyl muramoyl-glycyl-D- glutamate (UMGG) and its product UDP-N-acetyl muramoyl-glycyl-D-glu-*meso*-A₂pm (UTP) and also with ATP. The docked complexes reveal the amino acids responsible for binding the substrates. Superposition of these complex structures suggests that carboxylic acid group of UMGG is positioned in proximity to γ -phosphate of the ATP to facilitate the

formation of acylphosphate intermediate. The orientation of an amino group of A₂pm facilitates the nucleophilic attack to form the product. Overall, the proposed model together with its binding features gained from docking studies could help to design a truly selective ligand inhibitor specific to MurE for the treatment of leprosy.

Keywords Comparative modeling · Docking · Leprosy · *M. leprae* · Peptidoglycan

Introduction

Leprosy is a chronic infectious disease caused by *Mycobacterium leprae*, an obligate intracellular microorganism. The World Health Organization (WHO) estimated that 2 million people worldwide were infected with *Mycobacterium leprae* [1].

The comparative genome analysis of *M. leprae* (3.27 Mb) with other mycobacterium reveals a drastic gene reduction of *M. leprae* genome [2]. As *M. leprae* compensates the drastic gene reduction by a host-dependent parasitic life style, multiplication of *M. leprae* has not been achieved in bacteriological media or in macrophages infected with host derived *M. leprae* [3]. Hence the bioinformatics approaches will be helpful for the investigators in leprosy research.

Though the combination of drugs Dapsone, Rifampin and Clofazimine proved to be effective [4–9] for the treatment of leprosy, while their long term treatment led to mutations in *folP1*, *rpoB*, and *gyrA* genes which leads to the emergence of multidrug-resistant *M. leprae* [10]. Much of this resistance is attributable to the cell wall. Mycobacterium

A. Shanmugam
Department of Bioinformatics, VMKV Engineering College,
Vinayaka Missions University,
Salem 636 308, India
e-mail: anusuya_bioin@yahoo.co.in

J. Natarajan (✉)
Department of Bioinformatics, Bharathiar University,
Coimbatore 641 046, India
e-mail: n.jeyakumar@yahoo.co.in

has an unusually impermeable cell wall that contributes to their considerable resistance to many drugs.

The mycobacterial cell wall has unique characteristics and is impermeable to a number of compounds, a feature in part responsible for inherent resistance to numerous drugs. The mycobacterial cell wall consists of an inner layer and an outer layer that surround the plasma membrane. The outer compartment consists of both lipids and proteins. The inner compartment consists of peptidoglycan (PG), arabinogalactan (AG), and mycolic acids (MA) covalently linked together to form a complex (MAPc) known as the MA-AG-PG complex. This complex is insoluble and referred to as the essential core of the mycobacterial cell wall [11]. Many of the drugs used to combat mycobacteria target the MA-AG-PG complex [12].

Peptidoglycan or murein is a versatile material, rigid enough to provide a scaffold for bacteria to maintain their shape and protect them from osmotic pressure yet malleable enough for the bacteria to grow and expand [12]. PG is made of peptides and glycan strands. The long glycan strand typically consists of repeating *N*-acetylglucosamines (NAGs) linked to *N*-acetylmuramic acid (NAM). These strands are cross-linked by peptides bound to the lactyl group on NAMs from different glycan strands [12].

Biosynthesis of peptidoglycan is a multi-step (11–12 steps) process comprising three main stages:

1. Formation of UDP-*N*-acetylmuramic acid (UNAM) from *N*-acetylglucosamine (GlcNAc)
2. Addition of a short polypeptide chain to UNAM
3. Addition of a second GlcNAc to the disaccharide-pentapeptide building block and transport of this unit through the cytoplasmic membrane and incorporation into the growing peptidoglycan layer.

Stage two involves four key Mur ligase enzymes: MurC (EC:6.3.2.8), MurD (EC:6.3.2.9), MurE (EC:6.3.2.13) and MurF (EC:6.3.2.10). These four Mur ligases are responsible for the successive additions of Glycine [13], D-Glutamate, meso-diaminopimelate (meso-A2pm) and a dipeptide D-alanyl-D-alanine to UNAM. Our recent work on computational genome analyses of metabolic enzymes in *M. leprae* identified these MurC, MurD, MurE and MurF enzymes as potential drug targets [14]. As all these enzymes are essential for cell viability, they are attractive targets for antibacterial chemotherapy.

UDP-*N*-acetylmuramoyl-glycyl-D-glutamate:meso-diaminopimelate ligase (MurE) is a cytoplasmic enzyme that catalyzes the addition of meso-diaminopimelic acid to nucleotide precursor UDP-*N*-acetylmuramoyl-glycyl-D-glutamate in the biosynthesis of bacterial cell-wall peptidoglycan [15]. MurE catalyzes the addition of the third amino acid residue of the peptide chain of peptidoglycan. This residue, generally a diamino acid, varies among the bacterial species:

meso-diaminopimelic acid (meso-A2pm) for most Gram-negative bacteria, L-lysine for Gram-positive bacteria, L-ornithine, meso-lanthionine, LL-diaminopimelic acid, L-diaminobutyric acid or L-homoserine [16]. As the structure of PG is unique to bacteria and the specificity of MurE differs with bacteria, this MurE enzyme may be an excellent target for therapeutics.

Knowledge on the three-dimensional structure of MurE enzyme will provide clue on its potentially important residues for substrate binding and/or catalytic mechanism and lead to structure based drug discovery [17]. In the present study, we modeled this protein by comparative modeling method using the X-ray crystal structure of MurE from *E. coli* (PDB: 1E8C) as a template. Docking analysis of *M. leprae* MurE model with its substrates enabled us to identify specific residues within the MurE binding pocket responsible for binding to substrates. This observation underlines the important residues should be considered while designing a better drug for multidrug resistant *M. leprae*.

Methods

The amino acid sequence of the MurE enzyme of *M. leprae* (530 residues long) was obtained from GenBank at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>). The GenBank accession number for *M. leprae* MurE is NP_301692. The amino acid sequence was subjected to BLASTP analysis against *H. sapiens* to identify the homologous sequences. The amino acid sequence was again subjected to BLASTP [18] analysis against PDB (<http://www.pdb.org/pdb/home/home.do>) [19] and fold recognition servers PHYRE (<http://www.sbg.bio.ic.ac.uk/~phyre/>) [20] and 3D-PSSM (<http://www.sbg.bio.ic.ac.uk/3dpssm/>) [21] separately to identify the suitable template for comparative modeling.

As the comparative modeling relies on a sequence alignment between the target sequence and the template sequence whose structure has been experimentally determined, the target and the template sequences were aligned using ClustalW at the European Bioinformatics Institute (<http://www2.ebi.ac.uk/clustalw/>) [22].

After careful examination of the potential alignment errors, the automated comparative protein modeling program MODELLER9v6 was used to build the model [23]. In the first step of model building, distance and dihedral angle restraints on the target sequence were derived from its alignment with the template 3D-structure. The spatial restraints and the energy minimization steps were performed with the CHARMM22 force field for proper stereochemistry of proteins. Then, optimization of the model was carried out by the molecular dynamics simulated annealing method.

The *M. leprae* MurE model generated was evaluated using NIH SAVES metasever (<http://nihserver.mbi.ucla.edu/SAVES/>) [24]. The quality of the model is also assessed by comparing predicted structure to X-ray solved structure via superposition and measurement of the C- α Root Mean Square Deviation assessment (RMSD). The predicted model and the template 3D structure were submitted to the server SUPERPOSE [25] (<http://wishart.biology.ualberta.ca/SuperPose/>) and the RMSD between predicted model and template was identified.

Validity of the structural model was assessed by docking with its substrates meso-diaminopimelic acid (A2pm) and UDP-N-acetylmuramoyl-glycyl-D-glutamate (UMGG) and its product UDP-N-acetylmuramoyl-glycyl-D-glu-meso-A2pm (UTP) and also with ATP. The structure of the substrates UMGG and meso-A₂pm and the product UMT were drawn using ACD/chemsketch [26] and their Simplified Molecular Input Line Entry System (SMILES) notation were obtained. The PDB file of the substrates and products were generated using online server “Online Smiles Translator and Structure File Generator” [27]. The PDB file of ATP was obtained from drugbank [28].

To find the binding affinities between *M. leprae* MurE and its substrates, an automated flexible docking of ligands at the active site of an enzyme was carried out using AutoDock 4.0 [29]. The grid maps representing the protein in the actual docking process were calculated with the aid of AutoGrid. The dimension of the grid was 40X40X40 points in each dimension for A2pm and 60X60X60 points in each dimension for ATP and UMGG with spacing of 0.375 Å between the grid points. Gasteiger charges were computed using AutoDock tools (ADT) on the atoms for each ligand. The AUTOTORS utility, included in the

AutoDock software, was used to define all possible torsions of ligand molecules for the docking algorithm. Docking parameters were as follows: ten docking trials, population size of 150, maximum number of energy evaluation ranges of 250000, maximum number of generations of 27,000, mutation rate of 0.02, cross-over rate of 0.8 and an elitism value of 1. Each job consisted of 100 independent runs. Other docking parameters were set to the software’s default values. Finally, resulting docking orientations lying within 1.5 Å in the root-mean square deviation (rmsd) tolerance of each other were clustered together and represented by the result with the most favorable free energy of binding (ΔG_b).

Results and discussion

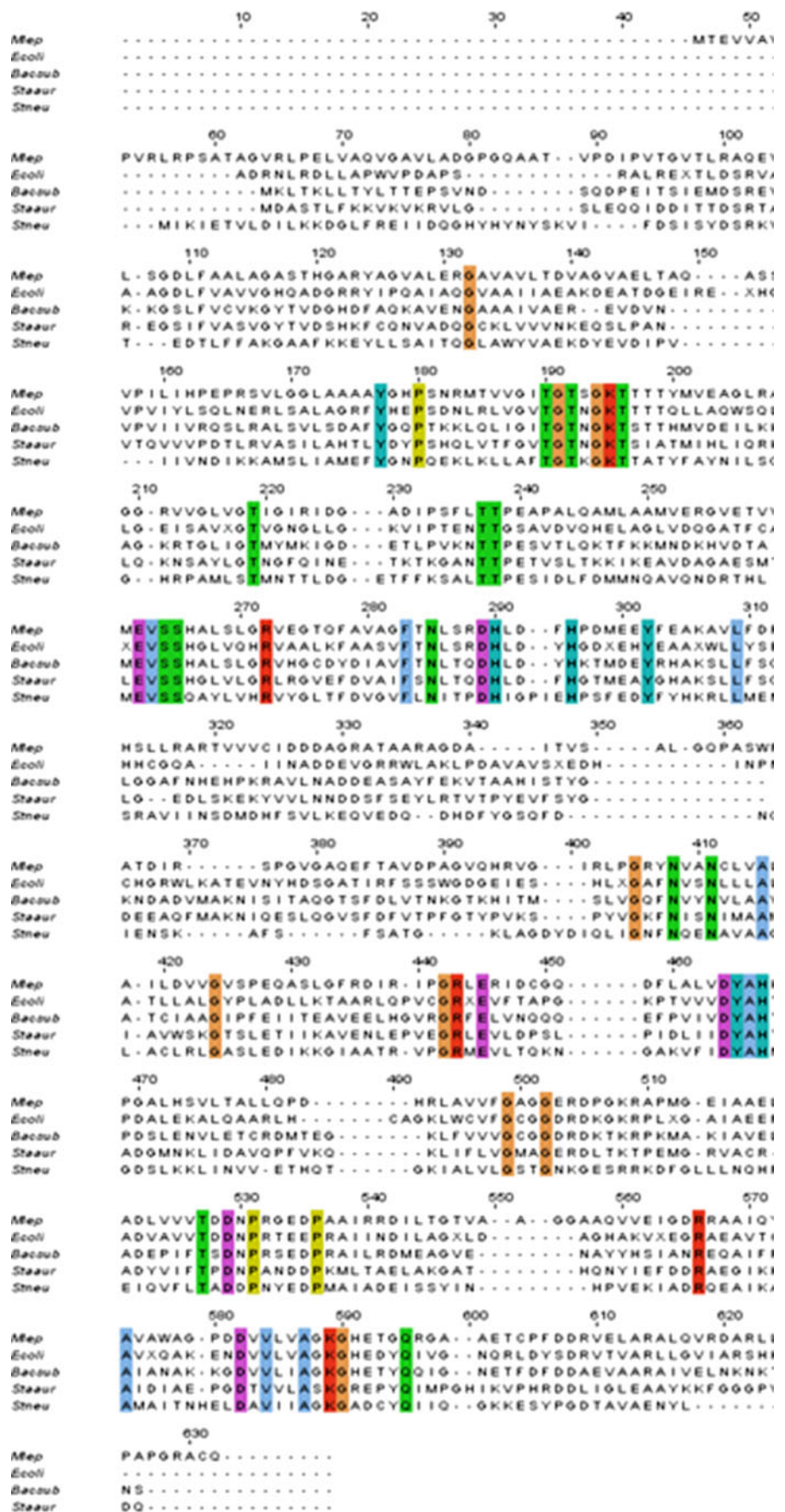
The amino acid sequence of *M. leprae* MurE was retrieved from GenBank database. Homologous sequences were not identified in the BLASTP search of *M. leprae* MurE against the *H. sapiens*. It suggested that MurE can be selected as a target for the treatment of leprosy and this could increase the probability for not producing any undesired interactions.

The BLASTP search of *M. leprae* MurE, the target sequence against PDB (Protein Data Bank) resulted in MurE, the UDP-N-acetylmuramyl tripeptide synthetase from *E. coli* (PDB: 1E8C) [15] with sequence identity of 40% at an *E*-value cutoff of $7e-55$. This result was supported by the results of 3D-PSSM and PHYRE servers. These findings were illustrated in Table 1. Hence, the structure of MurE, UDP-*N*-acetyl muramyl tripeptide synthetase from *E. coli* was chosen as the most suitable template for comparative modeling.

Table 1 Results of BLASTP, PHYRE and 3D-PSSM

PDB ID	Name of the protein	Source	% Identity		
			BLASTP	PHYRE	3D-PSSM
1E8C	UDP-N-acetylmuramyl tripeptide synthetase	<i>Escherichia coli</i>	40%	34%	35%
1E0D	UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase	<i>Escherichia coli</i>	29%	–	–
1W78	FOLC in complex with DHPP and ADP	<i>Escherichia coli</i>	28%	19%	–
2JFF	MurD ligase in complex with D-Glu containing Sulfonamide inhibitor	<i>Escherichia coli</i>	28%	14%	–
2UAG	UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase	<i>Escherichia coli</i>	28%	–	24%
1GG4	UDP-N-acetylmuramoyl-tripeptide D-alanyl-D-alanine adding enzyme	<i>Escherichia coli</i>	26%	25%	25%
2F00	UDP-N-acetylmuramoyl:L-alanine ligase	<i>Escherichia coli</i>	–	17%	–
1J6U	UDP-N-acetylmuramate-alanine ligase	<i>Thermotoga maritima</i>	26%	14%	16%
1GQQ	APO-Enzyme	<i>Haemophilus influenzae</i>	24%	–	22%
1P31	UDP-N-acetylmuramic acid:L-alanine ligase	<i>Haemophilus influenzae</i>	24%	17%	–
2AM1	UDP-N-acetylmuramoylalanine-D-glutamyl-lysine-D-alanyl-D-alanine ligase	<i>Streptococcus pneumoniae</i>	24%	19%	–

Fig. 1 Multiple sequence alignment of MurE ligases of five species



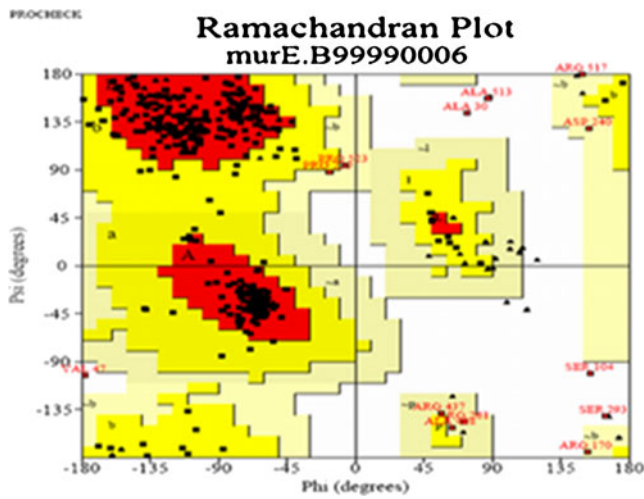


Fig. 2 Ramachandran plot of the *M. leprae* MurE model

The multiple sequence alignment of the target sequence with the homologous sequences obtained by the BLASTP was obtained using ClustalW. Seventy seven amino acids were found to be conserved among 70 MurE enzymes. Twelve of these conserved amino acids G142, K143, T144, T145, T166, T181, E206, S209, H210, R363 and R410 were found in active site region interact with the substrate and product. Hence they may determine catalytic activity of the enzyme. The remaining conserved amino acids may determine the structural properties of the protein. The multiple sequence alignment was shown in Fig. 1. For clarity, only five representative MurE ligase sequences were used (*M. leprae*, *E. coli*, *B. subtilis*, *S. aureus*, and *S. pneumoniae*).

Coordinates from the reference protein (PDB: 1E8C) for the structurally conserved regions (SCRs), structurally

Fig. 3 (a) Superposition of *M. leprae* MurE and X-ray crystal structure of *E. coli* MurE (1E8C) The predicted model of *M. leprae* MurE (in red color) superposed onto the structure of *E.coli* MurE (in yellow color) was shown in ribbon representation. (b) Superposition of *M. leprae* MurE and X-ray crystal structure of *M. tuberculosis* MurE (2WTZ) The predicted model of *M. leprae* MurE (in red) superposed onto the structure of and *M. tuberculosis* MurE (in yellow color) was shown in ribbon representation

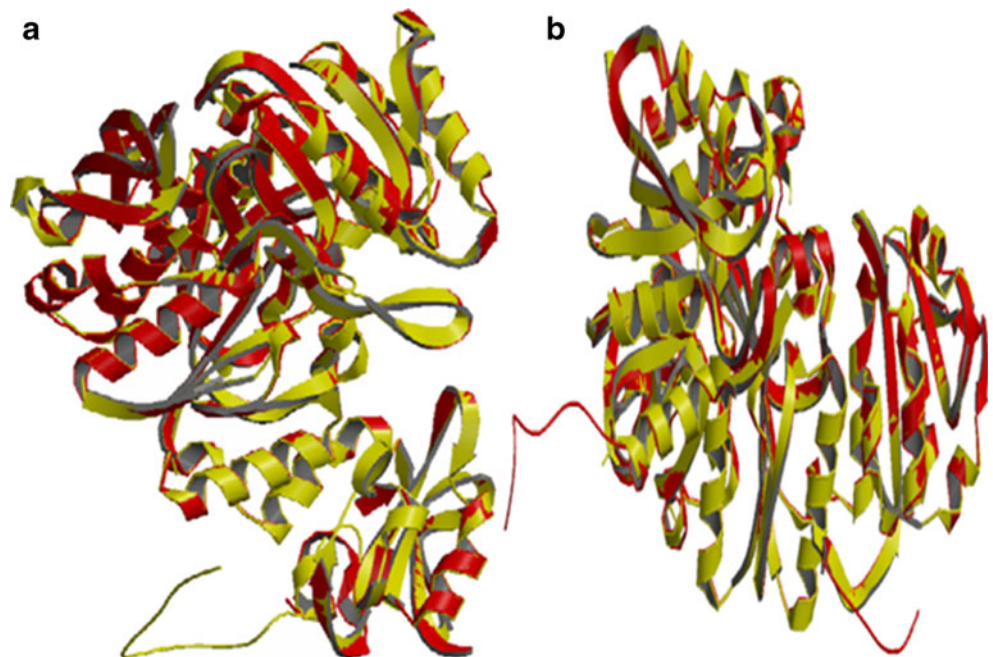


Fig. 4 Domains of *M. leprae* MurE in cartoon representation along with its substrates UMGG, ATP and meso-A2pm produced with Pymol. UDP binding domain is shown in cyan color, ATP binding domain in green color and ligand binding domain in purple blue color. The substrate UMGG is in black, ATP is in red and meso-A2pm is in blue color

variable regions (SVRs), N-terminal and C-terminal were assigned to the target sequence based on the satisfaction of spatial restraints. The initial model was generated using MODELLER. All side chains of the modeled protein were set by rotamers. The obtained model was refined by energy minimization using CHARMM force field and the final stable structure was obtained.

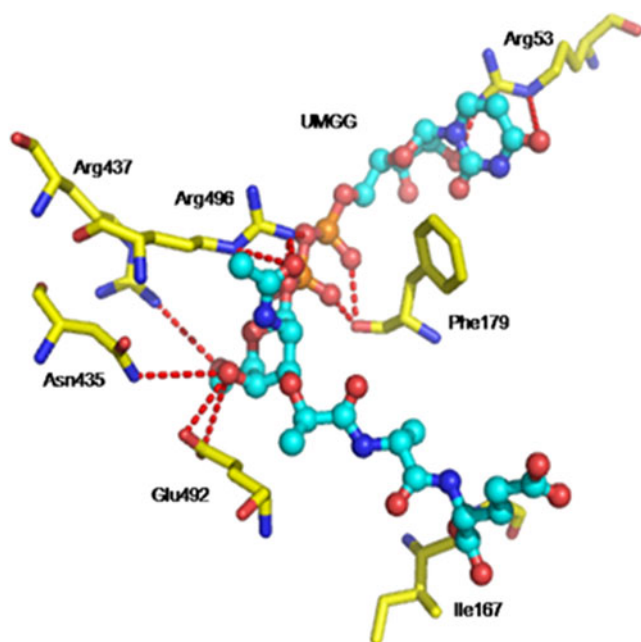


Fig. 5 MurE in complex with UMGG. Amino acids interact with UMGG were shown in stick representation in yellow color and the substrate UMGG was shown in ball and stick representation in cyan color using Pymol. The nitrogen, oxygen and phosphorous atoms were shown in blue, red and orange colors respectively. The hydrogen bonding interactions were shown in dotted lines

The selected model was subjected to “internal” evaluation of self-consistency checks such as stereo chemical check to find the deviations from normal bond lengths, dihedrals and non-bonded atom-atom distances. The PROCHECK Ramachandran plot analysis shows 91.7% of the residues in the most favoured region, 4.9% of residues in the additionally allowed region, 2.3% of residues in the generously allowed regions and 1.1% of residues in the disallowed region. His398, Asp397, Ala100, Pro326, Ser104 and Ser293 were found in the disallowed region in the Ramachandran plot. As these residues as well as their equivalent residues in *E.coli* were located too far from the

binding pocket, the model was considered for further studies. The Ramachandran plot for protein structure predicted is displayed in Fig. 2.

The goodness factors (G-factors) from the PROCHECK results summary confirmed the quality of covalent and overall bond/angle distances. The observed G-factors for the present model were -0.06 for dihedrals, -0.32 for covalent and 0.15 for overall. The comparable Ramachandran plot characteristics and the goodness factors supported the quality of the modeled structure.

A model can be a ‘highly accurate’ model, if the model have $\leq 2 \text{ \AA}$ RMSD from the experimentally determined structure. Such highly accurate models could fit for drug designing experiments and binding affinity tests. While models having C- α RMSD above this threshold and $\leq 4 \text{ \AA}$ were termed as “reliable” models. Such reliable models could fit for designing mutagenesis experiments but not for drug designing and binding affinity tests [30].

The quality of the model was also supported by the results of the server SUPERPOSE. SUPERPOSE compared the predicted structure of *M. leprae* MurE to X-ray solved structure of MurE from *E. coli* via superposition and was shown in Fig. 3a. The root mean square deviation (RMSD) of the modeled protein from the experimentally determined structure of *E. coli* MurE (1E8C) was found to be 0.34. As both sequences belong to mur ligase family, their sequence identity (40%) and the RMSD values suggested that this model is valid. The quality of the model is also validated using the recently published structure of MurE ligase from *M. tuberculosis* (2WTZ) [31] and was shown in Fig 3b. Our model showed RMSD value of 0.5 from the experimentally determined structure of MurE ligase from *M. tuberculosis*. As both these sequence belongs to mycobacteria, this again showed that the model is highly accurate and could fit for drug designing studies.

The *M. leprae* (*Mlep*), *E. coli* (*Ecoli*) and *B. subtilis* (*Bacsub*) enzymes incorporate *meso*-A2pm into the peptidoglycan, whereas the *S. aureus* (*Staur*) and *S. pneumoniae*

Table 2 Hydrogen bonds involved in UMGG binding

MurE residue	UMGG residue	Atom	Distance (in Å)
ARG53.NH	Uridine	N(42)	2.111
ARG53.NH	Ribose	O(51)	2.427
ILE167.O	D-Glu acid	O(1)	1.969
ILE167.O	D-Glu acid	O(3)	2.995
PHE179.O	B-PO4	O(31)	2.116
ARG496.NE	β -PO4	O(29)	2.691
ARG496.NH	β -PO4	O(29)	2.049
ARG496.NH	N-acetyl part of muramic acid	O(58)	2.139
ARG496.NE	N-acetyl part of muramic acid	O(58)	2.958
GLU492.OE	Muramic acid	O(23)	2.909

Table 3 Hydrogen bonds involved in ATP binding

MurE residue	Atoms of ATP	Distance (in Å)	MurE residue	Atoms of ATP	Distance (in Å)
GLY142.N	O(5)	2.892	GLU206.OE(1422)	O(6)	2.308
GLY142.N	O(5)	2.672	GLU206.OE(1423)	O(6)	3.077
LYS143.N	O(5)	2.068	HIS210.ND	O(14)	2.774
THR144.OG	O(6)	2.03	HIS210.ND	O(15)	2.283
THR144.OG	O(6)	2.088	HIS210.ND	O(16)	2.036
THR144.N	O(5)	2.66	HIS210.ND	O(16)	3.18
THR144.N	O(6)	2.205	HIS234.ND	O(14)	2.806
THR166.OG	O(7)	2.952	HIS234.NE	O(14)	2.208
THR181.OG	O(8)	2.208	HIS234.ND	O(14)	3.004
THR181.OG	O(9)	2.969	HIS234.ND	O(15)	3.251
THR181.OG	O(13)	3.062	ARG363.NH	O(4)	3.123
THR181.OG	O(12)	3.126	ARG363.NE	N(19)	3.172
TYR379.OH	O(10)	3.291	TYR379.OH	O(9)	2.054

(*Steu*) enzymes incorporate an L-lysine. Conserved residues in these MurE ligases are colored based on its nature.

In Ramachandran plot, the white areas correspond to conformations where atoms in the polypeptide come closer than the sum of their VanderWaals radii. These regions are sterically disallowed for all amino acids except glycine, which is unique in that it lacks a side chain. The red regions correspond to the allowed regions namely the alpha-helical and beta-sheet conformations where there are no steric

clashes. The yellow areas show the partially allowed regions of left handed helix wherein the atoms are allowed to come a little closer together.

Overall protein structure

MurE consists of three evolutionarily conserved globular domains as shown in Fig. 4.

Domain 1 present in the N- terminal comprises residues 46–125. It is composed of four β - sheets which were

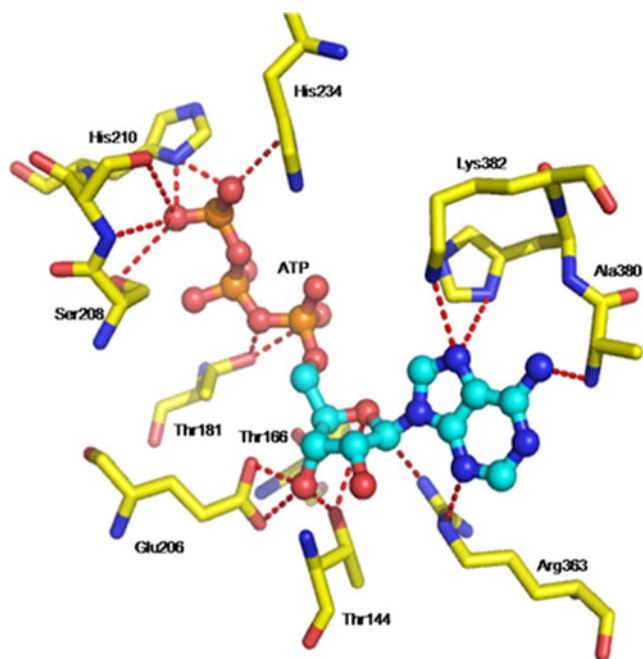


Fig. 6 MurE in complex with ATP. Amino acids interact with ATP was shown in stick representation in yellow color and the ATP molecule was shown in ball and stick representation in cyan color using Pymol. The nitrogen, oxygen and phosphorous atoms were shown in blue, red and orange colors respectively. The hydrogen bonding interactions were shown in dotted lines

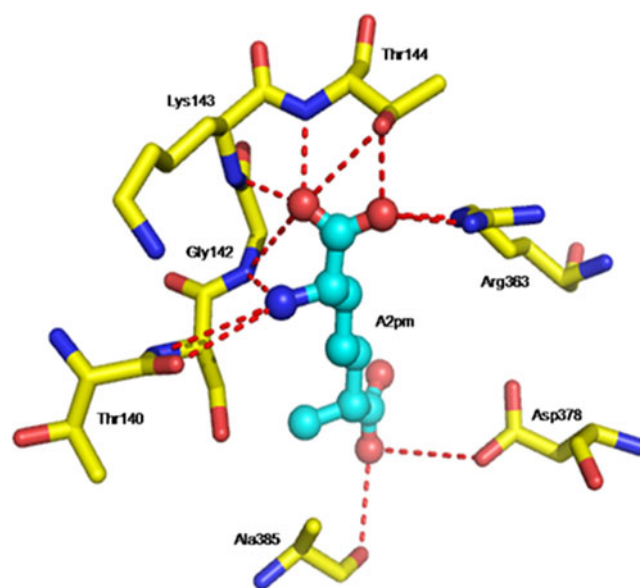


Fig. 7 MurE in complex with A₂pm. Amino acids interact with A₂pm was shown in stick representation in yellow color and A₂pm was shown in ball and stick representation in cyan color using Pymol. The nitrogen and oxygen atoms were shown in blue and red colors respectively. The hydrogen bonding interactions were shown in dotted lines

Table 4 Hydrogen bonds involved in meso-A₂pm binding

MurE residue	Meso-A ₂ pm residue	Atom	Distance (in Å)
GLY142.N	NH ₂	N(1)	2.403
LYS143.N	COOH	O(12)	2.508
THR144.N	COOH	O(12)	2.563
THR144.OG	COOH	O(13)	2.459
ARG363.NE	COOH	O(13)	2.447
ARG363.NH	COOH	O(13)	2.679
ALA385.O	COOH	O(10)	3.124
THR144.OG	COOH	O(12)	3.177

surrounded by three α -helices. Among the four β -sheets, three are found to be parallel to each other and one is anti parallel. This domain accounts for the fixation of UDP moiety of UMGG. This domain will therefore be referred to as UDP-binding domain.

Domain 2, the central domain comprises residues 135–341. This domain composed of two α -helices surrounded by ten β -sheets which were surrounded by five α -helices. Among the ten β -sheets seven are parallel to each other and three are anti parallel. The fold of the central β -sheet is similar to the classic “mononucleotide-binding fold” found in many ATP-binding proteins. This domain will therefore be referred to as the ATP binding domain.

Domain 3 present in the C- terminal comprises residues 360–445. This domain composed of four β -sheets surrounded by three α -helices. Among the four β -sheets three are found to be parallel to each other and one is anti parallel. This domain contains a classic Rossmann dinucleotide-binding fold, providing contacts to the ribose sugar and phosphate group. Residues from this ligand-binding domain provide key interactions that orient and position the incoming amino acid ligand with the growing peptidoglycan chain. This domain is also conserved among all of the Mur ligases.

Table 5 Hydrogen bonds involved in UMT binding

MurE residue	UMT residue	Atom	Distance (in Å)
ASP378.OD	Uridine	O8(8)	2.656
ASP378.OD	Uridine	N3(3)	2.945
LYS143.N	Ribose	O1(14)	2.626
THR144.N	Ribose	O1(14)	2.18
LYS143.N	Ribose	O1(15)	2.831
HIS234.NE	β -PO ₄	O2(23)	2.42
ARG210.NE	Muramic acid	O3(33)	2.392
HIS210.ND	Muramic acid	O3(35)	3.076
ALA69.N	D-Glu.acid	O5(54)	3.048

Binding site of UMGG

The substrate UMGG binds to MurE in the cleft between the three domains are shown in Fig. 5. The amino acids of MurE interacts with UMGG is tabulated in Table 2 Uridine part of UMGG is inserted between two loops of domain 1 connecting β 1 with β 2 and β 2 with α 1. The N of uracyl ring forms a hydrogen bond with side chain N of Arg53.

The pyrophosphate of UMGG interacts with the loop connecting β 8 with α 5 in ATP binding domain. The protein-phosphate interactions are made mainly through hydrogen bonds. One of the α -phosphate oxygen form hydrogen bond with Leu180. β -phosphate oxygen form hydrogen bonds with oxygen of Phe179, and N ϵ and NH of Arg496. The Arg496 is the only residue balances the negative charges of the pyrophosphate.

The *N*-acetylmuramic acid ring bridges the gap between the ATP-binding domain and domain 3. It is inserted between the loop connecting β 8 with α 5 of the ATP-binding domain and two of the loops of domain 3 connecting β 17 with α 12 and β 18 with α 13. Two groups of hydrogen bonds are important; O ϵ of Glu492 interact with O58, and Arg496.N ϵ and NH are in contact with the acetyl group.

Binding site of ATP

In nucleotide-binding proteins with the classical mononucleotide fold, there is a characteristic fingerprint, Gly-X-Gly-X-Gly-Lys-Thr/Ser located in the loop between a central β -strand and α -helix. A large anion hole is formed by the loop which accommodates the phosphates of the mono-nucleotide. In the MurE structure this loop comprises residues 137–144 with the sequence Gly-Ile-Thr-Gly-Thr-ser-Gly-Lys-Thr. It contains an additional residue between second and the third glycine residues of the motif and is located between β 6 and α 4 of the ATP binding domain. As this glycine rich G-loop holds the phosphates this may be called as P-loop.

Residues coordinating and stabilizing ATP include Gly142, Lys143, Thr144, Thr166, Thr181, Glu206, His210, His234 and Thr379. The amino acids of MurE interact with ATP by forming hydrogen bonds were given in Table 3 Lys143 and His234 balances the negative charges of the phosphates. Gly137, Gly142 and Lys143 in P-loop are conserved in all ADP forming proteins whereas the Gly140 was found to be conserved in all members of mur family enzymes. The amino acids forming hydrogen bond with ATP molecule is shown in Fig. 6.

Divalent cation is required for the activity of mur enzymes. Many of the enzymes containing the P-loop contain a divalent cation pocket, with one ligand supplied by the conserved hydroxyl residue of the P-loop and

Structural implications for catalysis

In general, ATP-dependent amide-forming enzymes are believed to share a common mechanism by catalyzing an initial phosphorylation of the acid carboxylate. Subsequently, the resulting acyl phosphate is attacked by the amine, producing a tetrahedral intermediate which ultimately collapses to the final product and inorganic phosphate [32] (Fig. 9). In order for the ligation to occur between UMGG and meso-A₂pm, the enzyme MurE must (i) bring together the UMGG and ATP, (ii) properly orient UMGG and ATP for the formation of an acyl phosphate intermediate, (iii) orient meso-A₂pm for the nucleophilic attack, and (iv) stabilize the tetrahedral intermediate, thereby lowering the activation barrier and accelerating catalysis.

The structure determination of MurE has revealed the location of the active site and identified the protein residues involved in the fixation of the substrate UMGG. In addition, the approximate location of the ATP-binding site has also been determined by structural homology with other ATP binding proteins. The active site of MurE is located in the cleft between the ATPase domain and domain 3. The reactive part of UMGG enters the cleft from the side closest to domain 1 and the ATP molecule from the opposite side. Residues involved in the fixation of UMGG include Arg53, Ile167, Phe179, Arg496 and Glu492.

The structure of *M. leprae* MurE in complex with UMGG, ATP and meso-A₂pm were superimposed and was shown in Fig. 4. The superimposed structure suggests that the carboxylate group of UMGG is in proximity with the γ -phosphate of ATP molecule to facilitate the formation of the acyl phosphate intermediate. The superimposed structure also suggests that the orientation of meso-A₂pm facilitates nucleophilic attack to form the UMT product.

ATP binding domain residue Ile167 is within hydrogen bonding distance of the carboxylate oxygen. Orientation and nearest-neighbor considerations suggest that the acyl phosphate is formed with UMGG carboxylate oxygen, implicating Ile167 as a likely participant in the formation of the acyl phosphate. Phe179 would also serve a role in orienting the carboxylate group prior to phosphorylation.

This shows that *M. leprae* MurE properly orient UMGG and ATP to facilitate the formation of acyl phosphate intermediate.

Comparison with *E. coli* MurE and *M. tuberculosis* MurE

The conserved amino acids important for the functionality in *M. leprae* were compared with that of MurE enzymes of *E. coli* and *M. tuberculosis* and shown in Table 6. The essential residues and their functionality for *E. coli* MurE and *M. tuberculosis* MurE were obtained from Gordan et al., [15] and Basavannacharya et al., [33] respectively. A striking detail in the comparison was the positional overlap between some of enzymatically important residues in all three ligases. In *M. tuberculosis* MurE D392, K157 (a residue in ATP binding consensus sequence GxxGKT/G) and E220 (a residue in second ATP binding motif) was predicted to be crucial for binding ATP [33]. Similar observations were also observed in *M. leprae* for ATP binding. The residue K143 present in the ATP binding consensus sequence and E206 present in second ATP binding motif were predicted to have a major role in binding ATP. In *M. tuberculosis* MurE R451 and N449 (residues of the DNPR motif) were found to be responsible for determining the specificity of meso-A₂pm incorporation into the peptidoglycan [33]. The same observation was also identified in *E. coli* MurE ligase [15]. In contrast to *M. tuberculosis* and *E. coli* murE ligases, a different binding pocket was observed for *M. leprae* murE for meso-A₂pm binding; In *M. leprae* R363 was found to have a crucial role in meso-A₂pm binding.

Conclusions

A molecular model of the MurE from *M. leprae* is documented in this study. The UDP binding domain, ATP binding domain and ligand binding domain were identified. The characteristic fingerprint for binding ATP molecule was identified. The evolutionarily conserved residues important

Table 6 Conserved amino acids essential for function in *M. leprae*, *M. tuberculosis* and *E. coli* MurE ligases

<i>M. leprae</i> MurE	<i>E. coli</i> MurE	<i>M. tub</i> MurE	Proposed Function in <i>M. lep</i> MurE
Lys143	Lys 119	Lys 157	ATP binding
Thr 181	Thr 157	Thr 195	ATP binding
His 234	His 210	His 248	UMT binding
Arg 363	–	–	UMT, A ₂ pm binding
Asp 378	Asp 356	Asp 392	UMT binding
His 381	His 359	–	UMT binding
Arg 410	Arg 389	–	UMT binding
–	Asn 449	Asn 414	A ₂ pm binding (in <i>E. coli</i> & <i>M. tuberculosis</i>)

for determining structural and functional aspects were found. Among the conserved amino acids G140, G142 and K143 played an important role in binding ATP, Glu206 was found to be important in fixing phosphates of ATP and in binding Mg^{2+} ion and Arg416 was found to be important in interacting with the free end of A_2pm . The superimposed structure of the docked complexes of MurE with its substrates and ATP and the orientation of amino group of meso- A_2pm suggests the nucleophilic attack to form the peptide bond. All important residues in *E. coli* MurD and MurE have corresponding residues in *M. leprae* MurE. This suggests that *M. leprae* MurE have the same catalytic machinery as that of the mur enzymes in *E. coli*. The present model and the binding features form the basis for further molecular studies on *M. leprae* MurE ligase inhibitor design for the treatment of multi drug resistant *M. leprae*.

References

- World Health Organization (2000) Leprosy-global situation. Wkly Epidemiol Rec 75:226–231
- Brosch R, Gordon SV, Eiglmeier K, Garnier T, Cole ST (2000) Comparative genomics of leprosy and tubercle bacilli. Res Microbiol 151:135–142
- Cole ST, Eiglmeier K, Parkhill J et al. (2001) Massive gene decay in the leprosy bacillus. Nature 409:1007–1011
- Levy L, Shepard CC, Fasal P (1976) The bactericidal effect of rifampicin on *M. leprae* in man: a) single doses of 600, 900 and 1200 mg; and b) daily doses of 300 mg. Int J Lepr Other Mycobact Dis 44:183–187
- Chemotherapy of leprosy. Report of a WHO study group. World Health Organ Tech Rep Ser 847:1–24 (1994)
- Norman G, Joseph G, Ebenezer G, Rao SP, Job CK (2003) Secondary rifampin resistance following multi-drug therapy—a case report. Int J Lepr Other Mycobact Dis 71:18–21
- Guelpa-Lauras CC, Grosset JH, Constant-Desportes M, Brucker G (1984) Nine cases of rifampin-resistant leprosy. Int J Lepr Other Mycobact Dis 52:101–102
- Ji BH (1985) Drug resistance in leprosy—a review. Lepr Rev 56:265–278
- Ji B (2002) Rifampin-resistant leprosy: a review and a research proposal of a pilot study. Lepr Rev 73:2–8
- Matsuoka M, Suzuki Y, Garcia IE, Fafutis-Morris M, Vargas-González A, Carreño-Martínez C, Fukushima Y, Nakajima C (2010) Possible mode of emergence for drug-resistant leprosy is revealed by an analysis of samples from Mexico. Jpn J Infect Dis 63:412–416
- Brennan PJ (2003) Structure, function and biogenesis of the cell wall of *Mycobacterium tuberculosis*. Tuberculosis (Edinb) 83:91–97
- Draper P, Kandler O, Darbre A (1987) Peptidoglycan and arabinogalactan of *Mycobacterium leprae*. J Gen Microbiol 133:1187–1194
- Mahapatra S, Crick DC, Brennan PJ (2000) Comparison of the UDP-*N*-acetylmuramate: L-alanine ligase enzymes from *Mycobacterium tuberculosis* and *Mycobacterium leprae*. J Bacteriol 182:6827–6830
- Shanmugam A, Natarajan J (2010) Computational genome analyses of metabolic enzymes in *Mycobacterium leprae* for drug target identification. Bioinformatics 4:392–395
- Gordon E, Flouret B, Chantalat L, van Heijenoort J, Mengin-Lecreulx D, Dideberg O (2001) Crystal structure of UDP-*N*-acetylmuramoyl-L-alanyl-D-glutamate: meso-diaminopimelate ligase from *Escherichia coli*. J Biol Chem 276:10999–11006
- Schleifer KH, Kandler O (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol Rev 36:407–477
- Gowthaman R, Silvester AJ, Saranya K, Kanya KS, Archana NR (2006) Modeling of the potential coiled-coil structure of snapin protein and its interaction with SNARE complex. Bioinformation 1:269–275
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402
- Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rogers JR, Kennard O, Shimanouchi T, Tasumi M (1978) The Protein Data Bank: a computer-based archival file for macromolecular structures. Arch Biochem Biophys 185:584–591
- Kelley LA, Sternberg MJ (2009) Protein structure prediction on the web: a case study using the Phyre server. Nat Protoc 4:363–371
- Jones DT, Taylor WR, Thornton JM (1992) A new approach to protein fold recognition. Nature 358:86–89
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) ClustalW and ClustalX version 2.0. Bioinformatics 23:2947–2948
- Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815
- Qadri YJ, Berdiev BK, Song Y, Lippton HL, Fuller CM, Benos DJ (2009) Psalmotoxin-1 docking to human acid-sensing ion channel-1. J Biol Chem 284:17625–17633
- Maiti R, Van Domselaar GH, Zhang H, Wishart DS (2004) SuperPose: a simple server for sophisticated structural superposition. Nucleic Acids Res 32 (Web Server issue):W590–594
- ACD/ChemSketch Freeware, version 10.00, Advanced Chemistry Development, Inc, Toronto, ON, Canada, www.acdlabs.com, 2006
- Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28:31–36
- Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 34 (Database issue): D668–672
- Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian Genetic algorithm and empirical binding free energy function. J Comput Chem 19:1639–1662
- Rayan A (2009) New tips for structure prediction by comparative modeling. Bioinformation 3:263–267
- Basavannacharya C, Robertson G, Munshi T, Keep NH, Bhakta S (2010) ATP-dependent MurE ligase in *Mycobacterium tuberculosis*: biochemical and structural characterization. Tuberculosis (Edinb) 90:16–24
- Bertrand JA, Auger G, Martin L, Fanchon E, Blanot D, Le Beller D, van Heijenoort J, Dideberg O (1999) Determination of the MurD mechanism through crystallographic analysis of enzyme complexes. J Mol Biol 289:579–590
- Basavannacharya C, Moody PR, Munshi T, Cronin N, Keep NH, Bhakta S (2010) Essential residues for the enzyme activity of ATP-dependent MurE ligase from *Mycobacterium tuberculosis*. Protein Cell 1:1011–1022

Interference of H-bonding and substituent effects in nitro- and hydroxy-substituted salicylaldehydes

Aneta Jezierska-Mazzarello · Halina Szatyłowicz ·
Tadeusz Marek Krygowski

Received: 10 January 2011 / Accepted: 16 March 2011 / Published online: 27 April 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Two intramolecular interactions, i.e., (1) hydrogen bond and (2) substituent effect, were analyzed and compared. For this purpose, the geometry of 4- and 5-X-substituted salicylaldehyde derivatives (X=NO₂, H or OH) was optimized by means of B3LYP/6-311+G(d,p) and MP2/aug-cc-pVDZ methods. The results obtained allowed us to show that substituents (NO₂ or OH) in the para or meta position with respect to either OH or CHO in H-bonded systems interact more strongly than in the case of di-substituted species: 4- and 3-nitrophenol or 4- and 3-hydroxybenzaldehyde by ~31%. The substituent effect due to the intramolecular charge transfer from the para-counter substituent (NO₂) to the proton-donating group (OH) is ~35% greater than for the interaction of para-OH with the proton-accepting group (CHO). The total energy of H-bonding for salicylaldehyde, and its derivatives, is composed of two contributions: ~80% from the energy of H-

bond formation and ~20% from the energy associated with reorganization of the electron structure of the systems in question.

Keywords Intramolecular hydrogen bond · H-bond energy · Homodesmotic reaction · DFT · MP2 · SESE

Introduction

Salicylaldehyde is a compound with well-recognized significance in many branches of chemistry. It undergoes a variety of chemical reactions, very often being a key precursor for new compounds exhibiting diverse molecular structures and properties [1–3]. It is worth mentioning that the salicylaldehyde moiety appears in many compounds exhibiting various biological activity, including reactants used in the design of new inhibitors of HIV-1 integrase [4], or compounds exhibiting antiviral activity [5], as well as in reactions resulting in new compounds with anticancer [6, 7] or antimicrobial activity [8]. It is also present during the synthesis of new products called “aspirin-like molecules” exhibiting anti-inflammatory activity [9]. In addition to its presence in many chemical reactions, salicylaldehyde has also found applications in molecular engineering [10].

Salicylaldehyde is also an interesting subject for various physico-chemical investigations; because it has four isomers, and due to internal reorganization, it can have more than one different hydrogen-bonded conformer [11]. The possible conformations of salicylaldehyde and hydrogen-bonded conformers are presented in Scheme 1.

The intramolecular hydrogen bond, spectroscopic signatures and geometric parameters of this compound have been of interest for the last several years. Many experimental as well as theoretical works on these issues have been

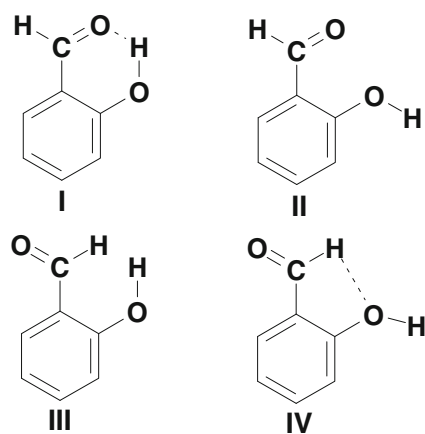
Dedicated to Professor Andrzej Górski, our friend and teacher, on the occasion of his 90th birthday.

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1044-1) contains supplementary material, which is available to authorized users.

A. Jezierska-Mazzarello (✉)
Faculty of Chemistry, University of Wrocław,
F. Joliot-Curie 14,
50-383 Wrocław, Poland
e-mail: anetka@elrond.chem.uni.wroc.pl

H. Szatyłowicz
Faculty of Chemistry, Warsaw University of Technology,
Noakowskiego 3,
00-664 Warsaw, Poland

T. M. Krygowski
Department of Chemistry, Warsaw University,
Pasteura 1,
02-093 Warsaw, Poland



Scheme 1 Possible conformations of salicylaldehyde (taken from [11]). Dashed line Intramolecular hydrogen bond

performed [12–19]. The intramolecular hydrogen bond present in salicylaldehyde can be classified as a resonance assisted hydrogen bond (RAHB) according to Gilli's concept [20]. The intramolecular hydrogen bond is of great importance in various aspects of chemistry, biology, and material science, as has been shown in selected examples [21–29]. An effort to understand and describe H-bonding can be observed in the literature, but there are still open questions related to the proton transfer phenomenon or the strength of the interaction, as well as environmental influences on it [30, 31]. The steric and inductive effects introduced by substituents can influence the strength of the intramolecular hydrogen bond significantly [32–40]. In the literature, one can find many papers dealing with the relationship between substituent effects and intramolecular H-bonding for acyclic systems [41, 42]. A very convenient system, malonaldehyde, has been used as a model for many studies [43–47]. One main conclusion from these studies is that the H-bond strength and π -electron delocalization in the OCCCO link in these systems depend on the type of substituent. All observed changes in the model systems are in agreement with Gilli's concept of RAHB [20, 48–50]; for further reading and review see [51, 52]. Concerning the intermolecular H-bond of para-substituted phenol/phenolate [53–55] or para-substituted aniline/anilide [56], the H-bonding and π -electron delocalization in the ring depend significantly on the kind of substituent.

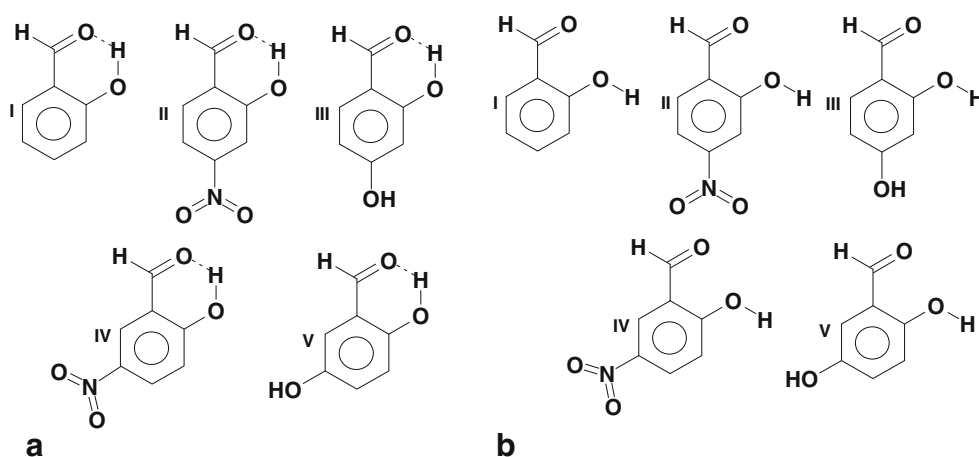
In the study, we focus on salicylaldehyde and its derivatives presented in Scheme 2. The choice of these compounds was governed by substituent character (electron accepting or donating) and its position with respect to CHO and OH groups; it is always a para position with respect to one of them. Salicylaldehyde was used mostly as a reference structure to investigate changes in the molecular structure upon benzene ring substitution as well as changes in the intramolecular hydrogen bond strength.

It is worth mentioning that estimating intramolecular H-bond energy is not an easy task due to the possible implications of internal reorganization of bonds or steric effects. Recent decades have seen increased interest in this problem. Various approaches have been introduced to handle the issue. Here, a few of these are briefly reported.

Let us start with a short overview of the simplest method based on conformational analysis—the cis-trans method. In this method, the intramolecular hydrogen bond energy is calculated as the difference between the cis and trans conformers—with and without interaction by H-bonding. Another approach is the so-called ortho-para method, which is restricted to aromatic compounds where the hydrogen bond is present in two ortho substituents [57]. Cuma et al. [16, 58] calculated the H-bond energy as the energy required to rotate the bridged hydrogen 180° from its equilibrium position around the appropriate C–O single bond. Based on this method, Grabowski [43] distinguished two components that comprise the strength of the intramolecular interaction: “pure” H-bond energy and delocalization energy. The so-called “theoretical” energy (ΔE_T) was calculated using a thermodynamic cycle, which describes the partition of the intramolecular hydrogen bond energy. This approach is described in [59]. Another briefly reported approach is based on the approximate isolation of the energy contribution that occurs upon the transition from one structure to other conformers of the studied molecule. The estimated energy is thus associated strictly with changes in geometric parameters (bonds and valence angles). The method is restricted to molecules with a suitable number of conformers of specific form [60]. Another way to estimate intramolecular H-bond energy is the “Molecular Tailoring Approach” proposed by Deshmukh et al. [61] This method is based on compound partitioning and energy estimation of appropriate fragments, which yields the intramolecular hydrogen bond energy. The last reported approach here is based on a hydrogen bond making/breaking reaction. Isodesmotic/homodesmotic reactions are written in such a way that the number and type of bonds on both sides (reactants/products) is equal [62, 63]. Therefore, it is possible to estimate the stabilization energy.

The main goal of the current study was to investigate how the substituent effect acts on π -electron delocalization and H-bond strength in the case of aromatic systems with intramolecular H-bonding, i.e., salicylaldehyde and its derivatives, based on energetic characteristics. Another problem discussed is associated with the fact that, as a result of intramolecular H-bond formation, the proton donating group increases its electron donating power, and the proton accepting group increases its electron accepting power [47, 64]. A question arises: how does this effect

Scheme 2a,b Structures of studied compounds: *I* Salicylaldehyde, *II* 4-nitro-salicylaldehyde, *III* 4-hydroxy-salicylaldehyde, *IV* 5-nitro-salicylaldehyde and *V* 5-hydroxy-salicylaldehyde respectively. **a** Salicylaldehyde and its derivatives containing the intramolecular hydrogen bonds (dashed lines), **b** open conformations

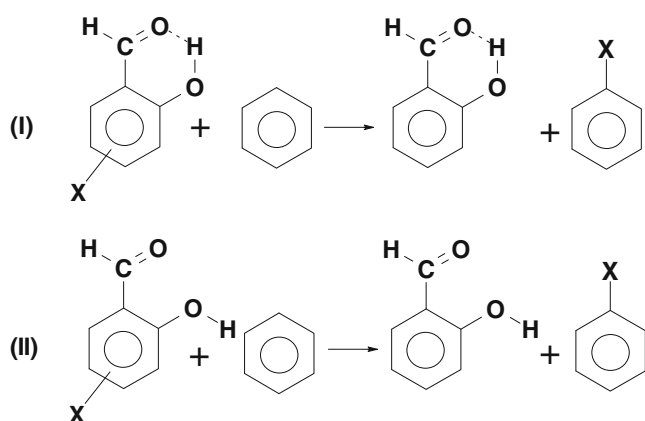


affect relations between substituent and H-bonding in the studied systems?

It should be stressed that both interactions, i.e., (1) H-bonding and (2) substituent effect, are intramolecular in nature. Therefore another question arises: do they cooperate or act in opposite directions?

Computational methodology

The energy minimization of the studied set of compounds (see Scheme 2a,b; and Schemes 3, 4, 5, 6 and 7 for mono- and di-substituted benzene derivatives) was performed using density functional theory (DFT) [65, 66] and the second-order Møller-Plesset perturbation (MP2) method [67]. For DFT calculations, 6-311+G(d,p) basis sets were used [68], whereas MP2 simulations were performed using the aug-cc-pVDZ basis set [69]. Concerning the DFT method, the three parameter hybrid functional proposed by Becke [70] with correlation energy according to the Lee-Yang-Parr formula [71], denoted as B3LYP, was employed.



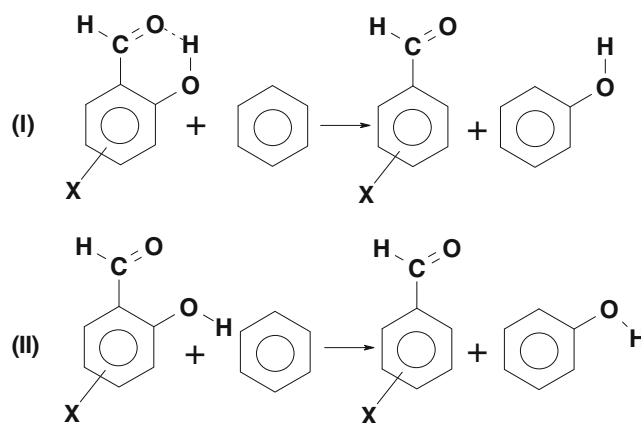
Scheme 3 Homodesmotic reactions (1) for closed and open conformations of 4- and 5-X substituted salicylaldehydes; X=NO₂ or OH

Subsequently, harmonic frequencies were calculated to confirm that the geometries obtained correspond to the minimum on the potential energy surface (PES). Next, single point calculations were performed to build up an additional set of close conformers, with the intramolecular hydrogen bond using open structures (without H-bond) and rotation of the O–H bond 180° around the C–O single bond.

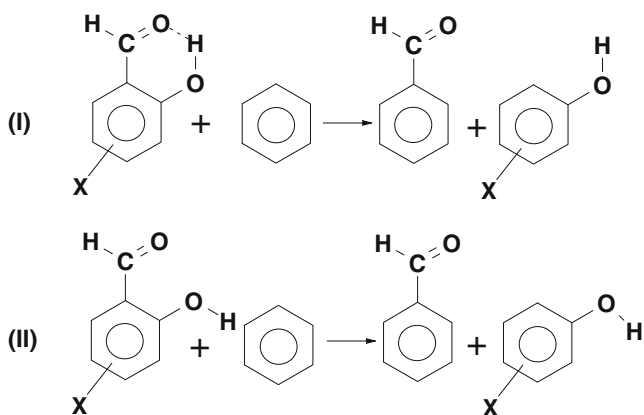
The estimation of the overall energy (E_{tot}) associated with the intramolecular H-bond formation was computed using the so-called cis-trans method, for details see Reference [72], as described below:

$$E_{\text{tot}} = E(\text{closed conformation}) - E(\text{open conformation}) \quad (1)$$

where E_{tot} indicates the total energy of the intramolecular hydrogen bond, whereas $E(\text{closed conformation})$ and $E(\text{open conformation})$ are the energies obtained after the geometry optimization procedures for closed and open conformations.



Scheme 4 Homodesmotic reaction (2) schemes for closed and open conformations of 4- and 5-X substituted salicylaldehydes; X=NO₂, H or OH



Scheme 5 Homodesmotic reaction (3) schemes for 4- and 5-X substituted salicylaldehydes; X=NO₂, H or OH

Two homodesmotic reaction schemes presented in Schemes 4 or 5 allow us to estimate the overall energy associated with intramolecular H-bond formation:

$$E_{\text{tot}} = \text{SESE}(\text{II}) - \text{SESE}(\text{I}) \quad (2)$$

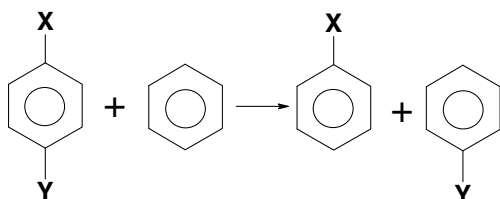
It is noteworthy that the final expression of the total energy, E_{tot} , could be transformed into the form of Eq. 1. In the case of substituted derivatives of salicylaldehyde, the two homodesmotic reaction schemes presented in Schemes 4 and 5 can be proposed. These reactions differ in products only, but this leads to a change of meaning in their substituent effect stabilization energy (SESE) value.

Following Grabowski's method [43, 45], the "pure" energy of H-bonding, E_{HB} , can be obtained as:

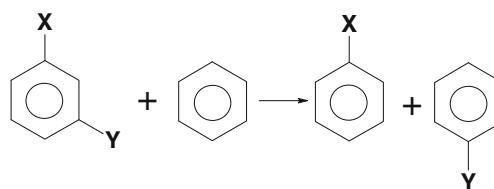
$$E_{\text{HB}} = E(\text{open conformation_O-H rotated } 180^\circ) - E(\text{open conformation, optimized}) \quad (3)$$

where $E(\text{open conformation_O-H rotated } 180^\circ)$ means the energy of the single point calculation for the closed conformer formed from that obtained from the open one after the rotation of O–H around the C–O single bond. The difference between the total and above energies is the energy due to the changes in geometry, sometimes called the energy of delocalization, E_{deloc} :

$$E_{\text{deloc}} = E_{\text{tot}} - E_{\text{HB}}. \quad (4)$$



Scheme 6 Homodesmotic reaction scheme for para di-substituted benzene derivatives; X=OH or CHO, Y=NO₂ or OH



Scheme 7 Homodesmotic reaction scheme for meta di-substituted benzene derivatives; X=OH or CHO, Y=NO₂ or OH

Furthermore, the SESE [63] was computed on the basis of the designed reactions in Schemes 3, 4, 5, 6 and 7. All computations were performed within the framework of the Gaussian03 and Gaussian09 suite of programs [73, 74]. The reaction schemes were prepared using the ISIS Draw v2.3 program [75].

Results and discussion

The mutual interference of substituent effects and the intramolecular hydrogen bond in a set of 4- and 5-substituted salicylaldehyde derivatives (see Scheme 2) was analyzed. The energetic characteristics concerning the coupling of the substituent effect and the presence of intramolecular hydrogen bonds was performed using two approaches considered at two levels of computations (B3LYP/6-311+G(d,p) and MP2/aug-cc-pVDZ):

- (1) Analysis of composition of the total energy of interaction, E_{tot} , for H-bonded systems perturbed by substituents.
- (2) Analysis of SESE for homodesmotic reactions for substituted salicylaldehyde and appropriate disubstituted benzene derivatives: 4-substituted benzaldehyde and 4-substituted phenols (substituent: OH and NO₂).

Two of the four possible conformations of salicylaldehyde (see Scheme 1) and its derivatives were considered in the study. Conformation (I) presented in Scheme 1 was found to be the most stable in many studies, e.g., [11], whereas the open conformation, labelled (II) in Scheme 1, was necessary for the analysis presented in this study.

Table 1 summarises all the data concerning the strength of intramolecular H-bonding, where the studied systems are presented in sequence from the strongest down to the weakest, taking into account the overall energy of the H-bond, E_{tot} . Almost the same order was found in the case of "pure" H-bonding energy, E_{HB} . The only difference with respect to the sequence used in Table 1 concerns salicylaldehyde and its 5-nitro derivative, but the difference in the E_{HB} value was very small (amounts to 0.03 kcal mol⁻¹). The same was found for E_{tot} , so one can draw the conclusion that the nitro group in the meta position with respect to the CHO acts similarly to hydrogen.

Table 1 Intramolecular hydrogen bond energy profile for salicylaldehyde and its derivatives. E_{tot} Overall energy of the H-bond, E_{HB} “pure” H-bonding energy, E_{deloc} energy of delocalization

Level of theory	E_{tot} kcal/mol	E_{HB} kcal/mol	E_{del} kcal/mol	$E_{\text{del}}/E_{\text{tot}}$ (%)
4-hydroxy-salicylaldehyde				
B3LYP/6-311+G(d,p)	-11.88	-9.14	-2.74	23.1
MP2/aug-cc-pvdz	-10.82	-8.55	-2.27	21.0
5-nitro-salicylaldehyde				
B3LYP/6-311+G(d,p)	-11.11	-8.91	-2.21	19.9
MP2/aug-cc-pvdz	-10.17	-8.28	-1.89	18.6
Salicylaldehyde				
B3LYP/6-311+G(d,p)	-11.08	-8.94	-2.15	19.4
MP2/aug-cc-pvdz	-10.20	-8.37	-1.83	17.9
5-hydroxy-salicylaldehyde				
B3LYP/6-311+G(d,p)	-10.82	-8.69	-2.13	19.7
MP2/aug-cc-pvdz	-10.02	-8.17	-1.84	18.4
4-nitro-salicylaldehyde				
B3LYP/6-311+G(d,p)	-10.32	-8.16	-2.16	20.9
MP2/aug-cc-pvdz	-9.79	-7.91	-1.87	19.1

Note that E_{tot} indicates as the most stable systems those in which a strong through-resonance effect is present: the para positions of CHO/OH and OH/NO₂ (i.e., 4-OH- or 5-NO₂-salicylaldehyde, respectively). These systems are more stable by ~1 kcal mol⁻¹ than complexes with substituents of the same kind: para positions of OH/OH and CHO/NO₂ (i.e., 5-OH- and 4-NO₂-salicylaldehyde, respectively). This kind of regularity is also observed for H-bond energies, E_{HB} , but the difference between those with and without through-resonance is smaller, ~0.6 kcal mol⁻¹. For the E_{del} the effect is even smaller, being equal to ~0.3 kcal mol⁻¹. The above mentioned data are from B3LYP/6-311+G(d,p) computations, but a similar picture

can be drawn from MP2/aug-cc-pVDZ calculations. It can be concluded from the data in Table 1 that the substituent effect on E_{tot} , E_{HB} and E_{del} is rather small. The variability of E_{HB} due to the substituent effect is also rather small (in the range of 0.98 kcal mol⁻¹), slightly smaller than that of E_{tot} (in the range of 1.56 kcal mol⁻¹), indicating resistance of H-bond interaction on perturbation stemming from substituents in the ring. Note that E_{del} , which is identified with changes in π -electron delocalization in the studied system, amounts to about 20% of E_{tot} .

A deeper insight into the intramolecular substituent effect gives an energetic characteristic of homodesmotic reactions (1), (2) and (3) (see Schemes 3, 4, 5, and Table 2).

Table 2 Substituent effect stabilization energy (SESE) of reactions (1), (2) and (3) calculated for salicylaldehyde and its 4-X- and 5-X- derivatives, X=NO₂ and OH; the homodesmotic reactions are shown in Schemes 3, 4, and 5, respectively. The sequence of the systems is the same as in Table 1 (from the strongest to the weakest in terms of H-bond overall energy)

Level of theory	SESE kcal/mol closed, I.1	SESE kcal/mol open, II.1	SESE kcal/mol closed, I.2	SESE kcal/mol open, II.2	SESE kcal/mol closed, I.3	SESE kcal/mol open, II.3
4-OH						
B3LYP/6-311+G(d,p)	2.37	1.58	8.25	-3.62	9.45	-2.43
MP2/aug-cc-pvdz	1.49	0.86	7.20	-3.62	8.11	-2.71
5-NO ₂						
B3LYP/6-311+G(d,p)	-0.56	-0.59	8.75	-2.36	5.55	-5.56
MP2/aug-cc-pvdz	-1.25	-1.22	7.15	-3.02	4.97	-5.20
H						
B3LYP/6-311+G(d,p)			7.36	-3.72		
MP2/aug-cc-pvdz			6.59	-3.61		
5-OH						
B3LYP/6-311+G(d,p)	-2.40	-2.14	5.33	-5.49	6.82	-4.00
MP2/aug-cc-pvdz	-1.75	-1.57	4.86	-5.16	6.48	-3.53
4-NO ₂						
B3LYP/6-311+G(d,p)	-3.33	-2.57	6.51	-3.81	4.18	-6.14
MP2/aug-cc-pvdz	-1.26	-0.85	6.70	-3.08	4.93	-4.86

Table 3 SESE calculated for para-di-substituted benzene derivatives. The homodesmotic reactions are illustrated in Scheme 6. Energies of the reaction components are given in Hartree

Level of theory					SESE (kcal/mol)
	4-NO ₂ -C ₆ H ₄ -OH + C ₆ H ₆ → C ₆ H ₅ -OH + C ₆ H ₅ -NO ₂				
	4-NO ₂ -C ₆ H ₄ -OH	Benzene	PhOH	PhNO ₂	
B3LYP/6-311+G(d,p)	-512.124008	-232.311245	-307.558632	-436.874621	1.25
MP2/aug-cc-pVDZ	-510.689883	-231.540220	-306.610690	-435.618828	0.37
	4-OH-C ₆ H ₄ -OH + C ₆ H ₆ → C ₆ H ₅ -OH + C ₆ H ₅ -OH				
	4-OH-C ₆ H ₄ -OH	Benzene	PhOH	PhOH	
B3LYP/6-311+G(d,p)	-382.803057	-232.311245	-307.558632	-307.558632	-1.86
MP2/aug-cc-pVDZ	-381.678535	-231.540220	-306.610690	-306.610690	-1.65
	4-NO ₂ -C ₆ H ₄ -CHO + C ₆ H ₆ → C ₆ H ₅ -CHO + C ₆ H ₅ -NO ₂				
	4-NO ₂ -C ₆ H ₄ -CHO	Benzene	PhCHO	PhNO ₂	
B3LYP/6-311+G(d,p)	-550.228511	-232.311245	-345.669087	-436.874621	-2.48
MP2/aug-cc-pVDZ	-548.680014	-231.540220	-344.603603	-435.618828	-1.38
	4-OH-C ₆ H ₄ -CHO + C ₆ H ₆ → C ₆ H ₅ -CHO + C ₆ H ₅ -OH				
	4-OH-C ₆ H ₄ -CHO	Benzene	PhCHO	PhOH	
B3LYP/6-311+G(d,p)	-420.918827	-232.311245	-345.669087	-307.558632	1.48
MP2/aug-cc-pVDZ	-419.675466	-231.540220	-344.603603	-306.610690	0.87

The obtained SESE data are specified in the same sequence as in Table 1. Positive values of SESE indicate greater stability of substrates, whereas negative values are products of the homodesmotic reaction.

The influence of the substituent on the formyl and hydroxyl groups—participants in the intramolecular hydrogen bond—is seen in the first reaction (**1**; see Scheme 3 and data in the second and third column in Table 2). The greatest strengthening of the H-bond compared with that in salicylaldehyde, is caused by the OH group in the para position with respect to CHO [4-OH-salicylaldehyde, SESE

(**1.1**) = 2.37 kcal mol⁻¹], whereas the opposite situation occurs in the case of the NO₂ group in the same position [4-NO₂-salicylaldehyde, SESE(**1.1**) = -3.33 kcal mol⁻¹]. A similar variability in SESE values is also observed for “open” conformations (1.58 and -2.57 kcal mol⁻¹, respectively), indicating a strong through-resonance effect in the first case and its absence in the second.

The ranges of SESE variability for reaction (**1**), calculated for 4-X- and 5-X- salicylaldehyde derivatives, X=NO₂ and OH, are 5.70 kcal mol⁻¹ and 4.15 kcal mol⁻¹ for the closed and open forms, respectively. This data can

Table 4 SESE calculated for meta-di-substituted benzene derivatives. Energies of the reaction components are given in Hartree

Level of theory					SESE (kcal/mol)
	3-NO ₂ -C ₆ H ₄ -OH + C ₆ H ₆ → C ₆ H ₅ -OH + C ₆ H ₅ -NO ₂				
	3-NO ₂ -C ₆ H ₄ -OH	Benzene	PhOH	PhNO ₂	
B3LYP/6-311+G(d,p)	-512.121766	-232.311245	-307.558632	-436.874621	-0.15
MP2/aug-cc-pVDZ	-510.689934	-231.540220	-306.610690	-435.618828	0.40
	3-OH-C ₆ H ₄ -OH + C ₆ H ₆ → C ₆ H ₅ -OH + C ₆ H ₅ -OH				
	3-OH-C ₆ H ₄ -OH	Benzene	PhOH	PhOH	
B3LYP/6-311+G(d,p)	-382.806466	-232.311245	-307.558632	-307.558632	0.28
MP2/aug-cc-pVDZ	-381.681107	-231.540220	-306.610690	-306.610690	-0.03
	3-NO ₂ -C ₆ H ₄ -CHO + C ₆ H ₆ → C ₆ H ₅ -CHO + C ₆ H ₅ -NO ₂				
	3-NO ₂ -C ₆ H ₄ -CHO	Benzene	PhCHO	PhNO ₂	
B3LYP/6-311+G(d,p)	-550.229361	-232.311245	-345.669087	-436.874621	-1.95
MP2/aug-cc-pVDZ	-548.679327	-231.540220	-344.603603	-435.618828	-1.81
	3-OH-C ₆ H ₄ -CHO + C ₆ H ₆ → C ₆ H ₅ -CHO + C ₆ H ₅ -OH				
	3-OH-C ₆ H ₄ -CHO	Benzene	PhCHO	PhOH	
B3LYP/6-311+G(d,p)	-420.915883	-232.311245	-345.669087	-307.558632	-0.37
MP2/aug-cc-pVDZ	-419.674037	-231.540220	-344.603603	-306.610690	-0.02

be compared with SESE values for para and meta disubstituted benzene derivatives; X=OH or CHO and Y=NO₂ or OH of the appropriate reactions (see Schemes 6, 7). The relevant data are presented in Tables 3 and 4. The greatest SESE value was found for 4-OH-C₆H₄-CHO (1.48 kcal mol⁻¹), supporting the strong through-resonance effect mentioned above, whereas the smallest was found for 4-NO₂-C₆H₄-CHO (-2.48 kcal mol⁻¹). The ranges of SESE variability for these cases are 3.96 kcal mol⁻¹ and 2.23 kcal mol⁻¹ for the para and meta substituted systems, respectively. Energetically, it means that the overall substituent effect on intramolecular H-bond in 4-X- and 5-X- substituted salicylaldehyde is ~31% larger than that observed in 3-X- and 4-X- substituted benzaldehyde or 3-X- and 4-X- substituted phenol. It should be noted that for the “open” conformations the effects compared above are similar to those observed for disubstituted benzene derivatives.

When we compare the ranges of SESE values for H-bonded systems (Table 2, 5.70 kcal mol⁻¹ for the closed system and 4.15 kcal mol⁻¹ for the open one), those for hydroxy- and nitro- benzaldehyde and phenol (3.96 kcal mol⁻¹ for para derivatives, Table 3; 2.23 kcal mol⁻¹ for meta systems, Table 4) with the range of E_{tot} and E_{HB} (1.56 and 0.98 kcal mol⁻¹, respectively, Table 1) we find immediately that the substituent effect, which is energetically substantial, acts very weakly on the total energy of H-bond formation as well as on the H-bond energy itself.

Two remaining homodesmotic reactions for salicylaldehyde and its derivatives (presented in Schemes 4, 5, and S1–S5 in Supporting Information, and data in Table 2) show another aspect of the substituent effect on H-bond formation. The range of SESE values for the I.2 reaction (closed) is 3.42 kcal mol⁻¹. In this case the substituent interacting via a through-resonance effect (OH group) is para in relation to the CHO group involved in H-bond formation. It may be compared with the range of SESE values for reaction I.3 (closed) which is 5.27 kcal mol⁻¹. Note that, in this case, the substituent interacting via a through-resonance effect (NO₂ group) is para with respect to the OH group involved in H-bond formation as a proton donating group. Thus, energetically, the substituent effect due to the intramolecular charge transfer from the para-counter substituent (NO₂) to the proton-donating group (OH) is ~35% greater than for the interaction of the para-OH with the proton-accepting group (CHO). This may suggest that, due to intramolecular charge transfer, the proton-donating component of H-bonding is more sensitive to the substituent effect than the proton-accepting one. A weaker H-bonding in 5-nitro-salicylaldehyde than in 4-hydroxy-salicylaldehyde (Table 1) explains the obtained SESE value for meta-nitrobenzaldehyde (-1.95 kcal mol⁻¹, Table 4).

Conclusions

In the case of intramolecular H-bond formation, the proton donating group (OH) increases its electron donating power, and the proton-accepting group (CHO) increases its electron accepting power [47, 57]. As a result, counter substituents (NO₂ or OH) in the para position to them in H-bonded 4- or 5- substituted salicylaldehyde interact more strongly (by ~31%) than in the case of 4-nitrophenol or 4-hydroxybenzaldehyde. Despite this strengthening, this perturbation acts weakly on the total energy of H-bond formation, E_{tot} , as well as the H-bond energy itself, E_{HB} .

The substituent effect due to the intramolecular charge transfer from the para-counter substituent (NO₂) to the proton-donating group (OH) is ~35% greater than for the interaction of para-OH with the proton-accepting group, CHO.

The total energy of the intramolecular H-bonded system, E_{tot} , contains ~20% of the energy associated with electron redistribution of the whole system, E_{del} .

Computations carried out using B3LYP/6-311+G(d,p) and MP2/aug-cc-pVDZ levels of theory lead to equivalent results.

Acknowledgments We warmly thank Dr. Jamieson Christie (University College, London) for his friendly support and help during the manuscript preparation. A.J.-M. gratefully acknowledges the Poznań Supercomputing and Networking Center and the Academic Computer Center CYFRONET-KRAKÓW (Grant: KBN/SGI/UWrocl/078/2001) for providing computer time and facilities, and thanks the Ministry of Science and Higher Education of Poland for supporting this work under the grant no. N N204 306137. H.S. and T.M. K. thank the Interdisciplinary Center for Mathematical and Computational Modeling (Warsaw, Poland) for computational facilities, and, in addition, H. S. would like to acknowledge the Warsaw University of Technology for financial support.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Williams JW, Sadle A (1940) The reactions of ketene with salicylaldehyde and p-hydroxybenzaldehyde. *J Am Chem Soc* 62:2801–2803
- Sacconi L, Bertini I (1966) Complexes of copper(II) with Schiff bases formed from salicylaldehydes and N-substituted ethylenediamines. *Inorg Chem* 5:1520–1522
- Sales ZS, Mani NS (2009) An efficient intramolecular 1,3-dipolar cycloaddition involving 2-(1,2-dichlorovinyl)oxy aryldiazomethanes: a one-pot synthesis of benzofuro-pyrazoles from salicylaldehydes. *J Org Chem* 74:891–894
- Zhao H, Neamati N, Sunder S, Hong H, Wang S, Milne GWA, Pommier Y, Burke TR Jr (1997) Hydrazide-containing inhibitors of HIV-1 integrase. *J Med Chem* 40:937–941
- Wang PH, Keck JG, Lien EJ, Lai MMC (1990) Design, synthesis, testing, and quantitative structure–activity relationship analysis of

- substituted salicylaldehyde schiff bases of 1-amino-3-hydroxy-guanidine tosylate as new antiviral agents against coronavirus. *J Med Chem* 33:608–614
- Chinigo GM, Paige M, Grindrod S, Hamel E, Dakshanamurthy S, Chruszcz M, Minor W, Brown ML (2008) Asymmetric synthesis of 2,3-dihydro-2-arylquinazolin-4-ones: methodology and application to a potent fluorescent tubulin inhibitor with anticancer activity. *J Med Chem* 51:4620–4631
 - Abd-Elzaher MM, Moustafa SA, Labib AA, Ali MM (2010) Synthesis, characterization, and anticancer properties of ferrocenyl complexes containing a salicylaldehyde moiety. *Monatsh Chem* 141:387–393
 - Prisakar VI, Tsapkov VI, Buracheeva SA, Byrke MS, Gulya AP (2005) Synthesis and antimicrobial activity of coordination compounds of copper with substituted salicylaldehyde thiosemicarbazones. *Pharm Chem J* 39:313–315
 - Lazzarato L, Donnola M, Rolando B, Marini E, Cena C, Coruzzi G, Guaita E, Morini G, Fruttero R, Gasco A, Biondi S, Ongini E (2008) Searching for new no-donor aspirin-like molecules: a new class of nitrooxy-acyl derivatives of salicylic acid. *J Med Chem* 51:1894–1903
 - Gothelf KV, Thomsen A, Nielsen M, Clo E, Brown RS (2004) Modular DNA-programmed assembly of linear and branched conjugated nanostructures. *J Am Chem Soc* 126:1044–1046
 - Chung G, Kwon O, Kwon Y (1998) Theoretical study on salicylaldehyde and 2-mercaptobenzaldehyde: intramolecular hydrogen bonding. *J Phys Chem A* 102:2381–2387
 - Catalán J, Toriblo F, Acuña AU (1982) Intramolecular hydrogen bonding and fluorescence of salicylaldehyde, salicylamide, and o-hydroxyacetophenone in gas and condensed phases. *J Phys Chem* 86:303–306
 - Schreiber V, Melikova S, Rutkowski K, Shchepkin D, Shurukhina A, Koll A (1996) Temperature dependence studies and model calculations of $\nu(\text{OH})$ and $\nu(\text{OD})$ band shapes of salicylaldehyde. *J Mol Struct* 381:141–148
 - Lampert H, Mikenda W, Karpfen A (1997) Molecular geometries and vibrational spectra of phenol, benzaldehyde, and salicylaldehyde: experimental versus quantum chemical data. *J Phys Chem A* 101:2254–2263
 - Filarowski A, Koll A, Sobczyk L (2009) Vibrational spectra of o-hydroxyphenyl Schiff bases and related compounds. *Curr Org Chem* 13:287–298
 - Cuma M, Scheiner S, Kar T (1999) Effect of adjoining aromatic ring upon excited state proton transfer, o-hydroxybenzaldehyde. *J Mol Struct THEOCHEM* 467:37–49
 - Graña AM, Ríos MA, Rodríguez J (1991) Ab Initio study of the structure and tautomerism of internally hydrogen-bonded aromatic carbonyls: salicylaldehyde and o-hydroxyacetophenone. *Struct Chem* 2:575–580
 - Kovács A, Szabó A, Hargittai I (2002) Structural characteristics of intramolecular hydrogen bonding in benzene derivatives. *Acc Chem Res* 35:887–894
 - Borisenko KB, Bock CW, Hargittai I (1996) Molecular geometry of benzaldehyde and salicylaldehyde: a gas-phase electron diffraction and ab initio molecular orbital investigation. *J Phys Chem* 100:7426–7434
 - Gilli G, Belluci F, Ferretti V, Bertolesi V (1989) Evidence for resonance-assisted hydrogen bonding from crystal-structure correlations on the enol form of the β -diketone fragment. *J Am Chem Soc* 111:1023–1028
 - Grabowski SJ (ed) (2006) Hydrogen bonding—new insights. Challenges and advances in computational chemistry and physics 3. Springer, Dordrecht
 - Sobczyk L, Grabowski SJ, Krygowski TM (2005) Interrelation between H-bond and Pi-electron delocalization. *Chem Rev* 105:3513–3560
 - Takahashi O, Kohno Y, Nishio M (2010) Relevance of weak hydrogen bonds in the conformation of organic compounds and bioconjugates: evidence from recent experimental data and high-level ab initio MO calculations. *Chem Rev* 110:6049–6076
 - Frey PA, Cleland WW (1998) Are there strong hydrogen bonds in aqueous solutions? *Bioorg Chem* 26:175–192
 - Gerlt JA, Kreevoy MM, Cleland WW, Frey PA (1997) Understanding enzymic catalysis: the importance of short, strong hydrogen bonds. *Chem Biol* 4:259–267
 - Desiraju GR (1989) Crystal engineering. The design of organic solids. Elsevier, Amsterdam
 - Jeffrey GA, Saenger W (1990) Hydrogen bonding in biological structures. Springer, Berlin
 - Desiraju GR (2002) Hydrogen bridges in crystal engineering: interactions without borders. *Acc Chem Res* 35:565–573
 - Gilli G, Gilli P (2009) The nature of the hydrogen bond: outline of the comprehensive hydrogen bond theory. Oxford University Press, Oxford
 - Jeziarska A, Panek JJ, Mazzarello R (2009) Structural and electronic structure differences due to the O–H \cdots O and O–H \cdots S bond formation in selected benzamide derivatives: a first-principles molecular dynamics study. *Theor Chem Acc* 124:319–330
 - Jeziarska A, Novič M, Panek JJ (2009) The nature of hydrogen bonding in selected hydrazide derivatives investigated via static models and Car-Parrinello molecular dynamics. *Pol J Chem* 83:799–819
 - Hammett LP (1970) Physical organic chemistry, 2nd edn. McGraw-Hill, New York
 - Exner O (1972) In: Chapman NB, Shorter J (eds) Advances in linear free energy relationships. Plenum, London
 - Johnson CD (1973) The Hammett equation. Cambridge University Press, Cambridge
 - Charton M (1981) Electrical effect substituent constants for correlation analysis. *Progr Phys Org Chem* 13:119–151
 - Exner O (1988) In: Chapman NB, Shorter J (eds) Correlation analysis of chemical data. Plenum, New York
 - Shorter J (1991) In: Zalewski RI, Krygowski TM, Shorter J (eds) Similarity models in organic chemistry, biochemistry and related fields. Elsevier, Amsterdam
 - Hansch C, Leo A, Taft RW (1991) A survey of hammett substituent constants and resonance and field parameters. *Chem Rev* 91:165–195
 - Exner O, Bohm S (2006) Theory of substituent effects: recent advances. *Curr Org Chem* 10:763–778
 - Krygowski TM, Stepień BT (2005) Sigma- and pi-electron delocalization: focus on substituent effects. *Chem Rev* 105:3482–3512
 - Krygowski TM, Szatyłowicz H (2006) Interrelation between the substituent effects, pi-electron delocalization and H-bonding. *Trends Org Chem* 11:37–53
 - Krygowski TM, Szatyłowicz H (2008) Modeling the Intermolecular H-bonded systems with varying energy of interactions: substituted phenols. In: Putz MV (ed) Advances in quantum chemical bonding structures. Transworld Research Network, Kerala, India, pp 287–308
 - Grabowski SJ (2001) An estimation of strength of intramolecular hydrogen bonds—ab initio and AIM studies. *J Mol Struct* 562:137–143
 - Grabowski SJ (2002) Properties of a ring critical point as measures of intramolecular H-bond strength. *Monatsh Chem* 133:1373–1380
 - Grabowski SJ (2005) π -Electron delocalisation for intramolecular resonance assisted hydrogen bonds. *J Phys Org Chem* 16:797–802
 - Krygowski TM, Zachara JE (2005) How pi-electron delocalisation reflects replacement of H $^+$ with Li $^+$ in variously substituted malonaldehydes. *Theor Chem Acc* 114:229–234

47. Krygowski TM, Zachara-Horeglad JE (2009) Resonance-assisted hydrogen bonding in terms of substituent effect. *Tetrahedron* 65:2010–2014
48. Bertolasi V, Gilli P, Ferretti V, Gilli G (1991) Evidence for resonance-assisted hydrogen-bonding. 2. Intercorrelation between crystal-structure and spectroscopic parameters in 8 intramolecularly hydrogen-bonded 1,3-diaryl-1,3-propanedione enols. *J Am Chem Soc* 113:4917–4925
49. Gilli P, Bertolasi V, Ferretti V, Gilli G (1994) Covalent nature of the strong homonuclear hydrogen-bond-study of the O–H–O system by crystal-structure correlation methods. *J Am Chem Soc* 116:909–915
50. Bertolasi V, Gilli P, Ferretti V, Gilli G (1996) Resonance-assisted O–H–O hydrogen bonding: Its role in the crystalline self-recognition of beta-diketone enols and its structural and IR characterization. *Chem Eur J* 2:925–934
51. Gilli G, Gilli P (2000) Towards an unified hydrogen-bond theory. *J Mol Struct* 552:1–15
52. Gilli P, Pretto L, Bertolasi V, Gilli G (2009) Predicting hydrogen-bond strengths from acid-base molecular properties. The pKa slide rule: toward the solution of a long-lasting problem. *Acc Chem Res* 42:33–44
53. Krygowski TM, Zachara JE, Szatyłowicz H (2004) Molecular geometry as a source of chemical information. 3. How H-bonding affects aromaticity of the ring in the case of phenol and p-nitrophenol complexes: A B3LYP/6-311+G** study. *J Org Chem* 69:7038–7043
54. Krygowski TM, Szatyłowicz H, Zachara JE (2004) How H-bonding affects aromaticity of the ring in variously substituted phenol complexes with bases. 4. Molecular geometry as a source of chemical information. *J Chem Inf Comput Sci* 44:2077–2082
55. Krygowski TM, Szatyłowicz H, Zachara JE (2005) Molecular geometry as a source of chemical information. 5. Substituent effect on proton transfer in para-substituted phenol complexes with fluorides at B3LYP/6-311+G**. *J Chem Inf Model* 45:652–656
56. Szatyłowicz H, Krygowski TM, Zachara-Horeglad JE (2007) Long-distance structural consequences of H-bonding. How H-bonding affects aromaticity of the ring in variously substituted aniline/anilinium/anilide complexes with bases and acids. *J Chem Inf Model* 47:875–886
57. Estácio SG, Cabral do Couto P, Costa Cabral BJ, Minas Da Piedade ME, Martinho Simões JA (2004) Energetics of intramolecular hydrogen bonding in di-substituted benzenes by the ortho-para method. *J Phys Chem A* 108:10834–10843
58. Scheiner S, Kar T, Čuma M (1997) Excited state intramolecular proton transfer in anionic analogues of malonaldehyde. *J Phys Chem A* 101:5901–5909
59. Lipkowski P, Koll A, Karpfen A, Wolschann P (2002) An approach to estimate the energy of the intramolecular hydrogen bond. *Chem Phys Lett* 360:256–263
60. Jabłoński M, Kaczmerek A, Sadlej AJ (2006) Estimates of the energy of intramolecular hydrogen bonds. *J Phys Chem A* 110:10890–10898
61. Deshmukh MM, Gadre SR, Bartolotti LJ (2006) Estimation of intramolecular hydrogen bond energy via molecular tailoring approach. *J Phys Chem A* 110:12519–12523
62. Pross A, Radom L, Taft RW (1980) Theoretical approach to substituent effects—phenols and phenoxide ions. *J Org Chem* 45:818–826
63. George P, Trachtmann M, Bock CW, Brett AM (1975) Alternative approach to problem of assessing stabilization energies in cyclic conjugated hydrocarbons. *Theor Chim Acta* 38:121–129
64. Krygowski TM, Zachara-Horeglad JE, Palusiak M (2010) H-bonding-assisted substituent effect. *J Org Chem* 75:4944–4949
65. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. *Phys Rev* 136:B864–B871
66. Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. *Phys Rev* 140:A1133–A1138
67. Møller C, Plesset MS (1934) Note on an approximation treatment for many-electron systems. *Phys Rev* 46:618–622
68. Krishnan R, Binkley JS, Seeger R, Pople JA (1980) Self-consistent molecular-orbital methods. 20. Basis set for correlated wave-functions. *J Chem Phys* 72:650–654
69. Dunning TH Jr (1989) Gaussian-basis sets for use in correlated molecular calculations. 1. The atoms boron through neon and hydrogen. *J Chem Phys* 90:1007–1023
70. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98:5648–5652
71. Lee C, Yang W, Parr RG (1988) Development of the correlation-energy formula into a functional of the electron density. *Phys Rev B Condens Matter Mater Phys* 37:785–789
72. Dietrich SW, Jorgensen EC, Kollman PA, Rothenberg S (1976) A theoretical study of intramolecular hydrogen bonding in ortho-substituted phenols and thiophenols. *J Am Chem Soc* 98:8310–8324
73. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) *Gaussian03, Rev C02*. Gaussian Inc, Wallingford CT
74. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam NJ, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) *Gaussian 09, Rev A1*. Gaussian Inc, Wallingford CT
75. *IsisDraw v2.3*. © 1990–2000 MDL Information System, Inc

DFT studies of the phenol adsorption on boron nitride sheets

Jose Mario Galicia Hernández ·
Gregorio Hernández Coccoletzi · Ernesto Chigo Anota

Received: 10 January 2011 / Accepted: 17 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract We perform first principles total energy calculations to investigate the atomic structures of the adsorption of phenol (C_6H_5OH) on hexagonal boron nitride (BN) sheets. Calculations are done within the density functional theory as implemented in the DMOL code. Electron-ion interactions are modeled according to the local-spin-density-approximation (LSDA) method with the Perdew-Wang parametrization. Our studies take into account the hexagonal h-BN sheets and the modified by defects d-BN sheets. The d-BN sheets are composed of one hexagon, three pentagons and three heptagons. Five different atomic structures are investigated: parallel to the sheet, perpendicular to the sheet at the B site, perpendicular to the sheet at the N site, perpendicular to the central hexagon and perpendicular to the B-N bond (bridge site). To determine the structural stability we apply the criteria of minimum energy and vibration frequency. After the structural relaxation phenol molecules adsorb on both h-BN and d-BN sheets. Results of the binding energies indicate that phenol is chemisorbed. The polarity of the system increases as a consequence of the defects presence which induces transformation from an ionic to covalent bonding. The elastic properties on the BN structure present similar behavior to those reported in the literature for graphene.

Keywords Boron nitride · DFT theory · Lattice defects · Phenol

Introduction

Boron nitride (BN) sheets were discovered for the first time by Geim et al. [1] in 2005. BN layers display hexagonal planar atomic structure similar to graphene [2]. Since the BN layer discovery these structures have received the attention of only a few works in the literature. However, recent theoretical studies have demonstrated that BN sheets may be used to adsorb molecules like ozone and water; these results suggest the possible use of BN layers as gas sensors [3, 4].

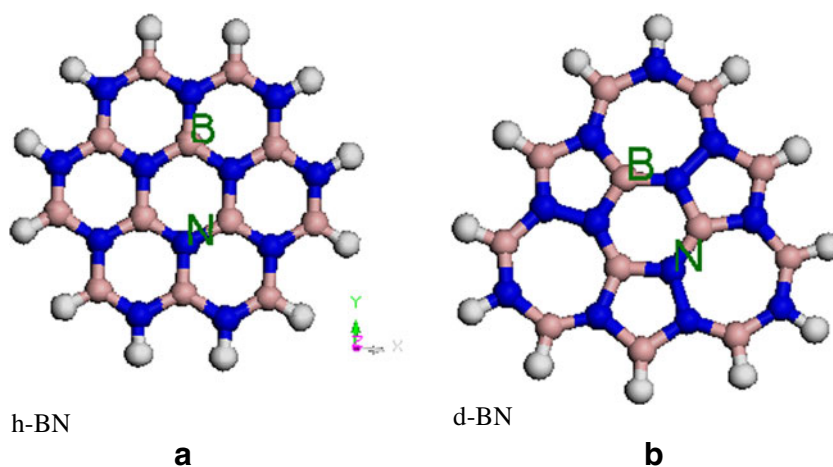
Phenol (C_6H_5OH), also known as phenyl acid or carboic acid [5], has a non-alcohol atomic structure as its functional group is Ph-OH, which is different from that of alcohol, R-OH. It is mainly used to produce resins, nylon, synthetic fibers, agrochemicals, polycarbonates and acetylsalicylic acid (aspirin). Phenol is employed in the preparation of orally cleaning liquids and pills to cure sore throat. It is also used as fungicide, antibacterial [6], antiseptic and disinfectant in the chemistry, pharmacy and clinic industries [7–9]. However, phenol is a contaminant in products like wastewater [10–12], therefore in this work we address the problem of studying its adsorption on the hexagonal h-BN (Fig. 1a) [3, 4] and on the modified by defects d-BN sheet structures (Fig. 1b) [13], in order to search for a material suitable to remove contaminants.

Atomic and electronic properties investigations of d-BN sheets have been motivated in part by the recent fabrication of graphene with lattice defects [14]. According to Martínez et al. [15] and You et al. [16] a d-BN layer may be modeled as an armchair edge. This model has been successfully used

J. M. Galicia Hernández · G. H. Coccoletzi
Benemérita Universidad Autónoma de Puebla, Instituto de Física
'Luis Rivera Terrazas',
Apartado Postal J-48,
Puebla 72570, Mexico

E. C. Anota (✉)
Benemérita Universidad Autónoma de Puebla, Facultad de
Ingeniería Química, Cuerpo Académico de Ingeniería en
Materiales, Ciudad Universitaria,
San Manuel,
Puebla, Código Postal 72570, Mexico
e-mail: echigoa@yahoo.es

Fig. 1 We depict models of the boron nitride sheets (a) hexagonal structure, h-BN, and (b) lattice modified by defects, d-BN [13]. These structures are on the xy-plane, atom colors are: Blue:N, Pink:B and White:H



to explain the doping of graphene with nitrogen [17], and the boron nitride with carbon [17], lithium and fluorine [18]. The model also helps to investigate the electronic properties of fluorine-graphene [19], the hexagonal structures of III-A group nitrides [20], the structural and chemical properties of graphene oxide [21] and to predict the formation of boron nitride oxide [22–24]. The atomic structure of d-BN sheets we consider in this work is composed of one hexagon, three pentagons and three heptagons and it is represented by a C_nH_m cluster.

Computational procedure

First principles total energy calculations are done, within the density functional theory [25–28] using the DMOL³ code [29], to investigate the adsorption of phenol on the hexagonal h-BN sheets and on the modified by defects d-BN sheets. The exchange-correlation energies are treated according to the local spin density approximation (LSDA) with the Perdew-Wang (PWC) parametrization [30], which accounts for the all-electron polarized atomic base (including a *p* orbital of H, and a *d* orbital of carbon, oxygen, boron and nitrogen) to deal with the core (DNP) [29, 31, 32]. Studies are performed with neutral ($Q=0$), cationic (+1) and anionic (−1) charges for the phenol–h-BN and phenol–d-BN systems. Considering both phenol–h-BN and phenol–d-BN configurations five atomic geometries have been analyzed for each structure; in the first array the phenol molecule is parallel to the central hexagon, and in the other four arrays the phenol molecule is placed perpendicular to the sheets: on top of the boron atom, on top of the nitrogen atom, above the central hexagon and above the B–N bond (see Fig. 2a–i). The hexagonal h-BN sheet has a diameter of 0.958 nm, and the d-BN structure has a base of 0.946 nm and a height of 0.986 nm. The orbital of the phenol–BN sheet is constructed with a cut radius of 0.41 nm and the phenol with a cut radius of 0.37 nm, with a tolerance of $1.0 \times$

10^{-6} Ha for the SCF cycle. To determine the structural stability we apply the criteria of minimum energy (see Table 1) and the non-negative vibration frequency [33]. Polarity (dipole moment), binding energies [for BN $E_b = (nE_B + mE_N + kE_H) - (E_{BN} + E_{ZPE})$ and for BN + Phenol $E_b = (nE_B + mE_N + kE_H + E_O) - (E_{BN} + E_{Phenol} + E_{ZPE})$] and adsorption energies [$E_{ad} = E_{total} - E_{Phenol} - E_{BN}$] are also obtained in the most stable geometries.

In order to validate the simulated models, the cohesion energy ($E_c = [nE(B) + mE(N) + kE(H) - E_{BNmHk} / (n + m + k)]$) of naphthalene-type ($B_5N_5H_8$), phirene type ($B_8N_8H_{10}$), coronene type ($B_{12}N_{12}H_{12}$), and the cluster $B_{27}N_{27}H_{18}$ were calculated, obtaining the value of 1.66 a.u./atom for all of the systems.

Results and discussion

This section is devoted to describing results for the adsorption of phenol on both the hexagonal h-BN (defects free) [34, 35] and the modified by defects d-BN structures. Calculations are presented for the total energy, polarity, adsorption energies and elastic properties.

Adsorption of phenol on the h-BN sheet

BN layers display hexagonal planar atomic structure similar to graphene [1]. To investigate the adsorption of phenol on the hexagonal h-BN sheets we consider different possible atomic orientations of the phenol molecule: Parallel and perpendicular to the h-BN structure. In the parallel configuration phenol is above the central hexagon (Fig. 2a). In the perpendicular configuration four arrays are explored: On top of the boron atom, on top of the nitrogen atom, above the central hexagon and above the B–N bond (see Fig. 2c, e, g and i).

Results of the total energy yield the parallel configuration as the one with the lowest minimum energy conse-

Table 1 We report the total energies for the geometries considered in the adsorption of phenol on boron nitride nanosheets

Configuration	Total Energy (a.u.)	Gap (HOMO-LUMO) in eV
h-BN + PHENOL		
C1 Parallel	-1260.4514133	4.09
C2 perpendicular to the hexagon	-1260.4481892	3.86
C3 perpendicular to B	-1260.4484198	4.14
C4 perpendicular to N	-1260.4435971	3.92
C5 perpendicular to the B-N bond	-1260.4421747	4.26
d-BN + PHENOL		
C1 Parallel	-1260.1426473	3.29
C2 perpendicular to the hexagon	-1260.1315860	3.21
C3 perpendicular to B	-1260.1355659	3.00
C4 perpendicular to N	-1260.1365823	2.98
C5 perpendicular to the B-N bond	-1260.1309815	3.27

quently this structure is considered as the ground state. The criteria of minimum energy and the non-negative vibration frequency are applied for the atomic relaxation. The perpendicular distance from the N atom of the h-BN sheet to the molecule is 3.13 Å (Fig. 3a). This is accompanied with the neutral charges ($Q=0$) and multiplicity 1. The B-N bond length is 1.44 Å. Taking into account the binding energy values and the criterion of positive vibration frequencies we can conclude that the sample synthesis is possible. A summary of parameters is presented in Table 2.

Concerning the polarity this is increased because of the phenol adsorption. Values go from 3.7×10^{-3} Debye to 1062.8×10^{-3} (neutral), to 862.3×10^{-3} (anionic) and to 1921.5×10^{-3} Debye (cationic). These results indicate that the adsorbed acid transfers charge to the h-BN sheets to induce a transition from an ionic (the dipole moment vector in the isolated sheet is directed toward the geometric center, Fig. 4b) to a covalent character (determined by the dipole moment vector, Fig. 4c). The value of the dipole moment indicates that the structure displays an ionic character in agreement with the experimental report given in Ref. [36]. The HOMO and LUMO orbitals exhibit an energy gap of 6.96 eV [13].

We have also studied charge effects on the atomic structure stability considering the following system in three different situations: (a) with neutral charge $[(C_6H_5OH)-h-BN]^0$, (b) with negative charge $[(C_6H_5OH)-h-BN]^{-1}$ (as anionic case) and (c) with positive charge $[(C_6H_5OH)-h-BN]^{+1}$ (as cationic case). Results indicate that it is possible to achieved stable atomic structures in these three configurations for the phenol adsorption on h-BN sheets. These atomic configurations exhibit semiconductor behavior under neutral (energy gap of 4.09 eV) and cationic (energy gap of 4.22 eV) charge conditions, and as a conductor under anionic charge, with a negligible energy gap of 5×10^{-2} eV. Calculations of the adsorption energies yield:

-0.47 eV for $[phenol-h-BN]^0$, -1.57 eV for $[phenol-h-BN]^{-1}$ and -1.48 eV for $[phenol-h-BN]^{+1}$. These results indicate that phenol is chemisorbed in all configurations.

The adsorption of phenol in the perpendicular geometries display interesting features, after relaxation the molecule plane is rotated with respect to the original vertical plane of the phenol: 30° for configuration 2 (Fig. 3c), 50° for configuration 3 (Fig. 3e) and 55° for configuration 4 (see Fig. 3g for atomic arrays). In configuration 5 (Fig. 3i) the phenol plane makes an angle of 15° with the vertical axis.

Adsorption of phenol on the d-BN sheet

The d-BN atomic array is produced by the presence of defects in the hexagonal h-BN structure. It is composed of one hexagon, three pentagons and three heptagons, and is represented by a C_nH_m cluster. Similar to the adsorption of phenol on the h-BN sheets, we explore different orientations of the phenol molecule: Parallel and perpendicular to the d-BN sheets. In the parallel geometry the molecule is above the central hexagon (Fig. 2b). In the perpendicular geometry four configurations are explored: On top of the boron atom, on top of the nitrogen atom, above the central hexagon and above the B-N bond (see Fig. 2d, f, h and j).

Similar to the adsorption of phenol on the h-BN sheets, results of the total energy yield the parallel configuration as the one with the lowest minimum energy (see Table 1) consequently this structure is considered as the ground state. The criteria of minimum energy and the non-negative vibration frequency [33] are again applied for the atomic relaxation. The perpendicular distance from the N atom of the d-BN sheet to the molecule is 3.52 Å (Fig. 3b). In this case there is neutral charge ($Q=0$) and multiplicity 1. The B-N bond length is 1.43 Å. Even though phenol is adsorbed above the central hexagon in an almost parallel geometry,

Fig. 2 These are the initial geometries for (a, b) parallel case, (c, d) perpendicular to the central hexagon, (e, f) perpendicular to the sheet at the N site, (g, h) perpendicular to the sheet at the B site, (i, j) perpendicular to sheet at the B-N bond (bridge). Atom colors are Blue: N, Pink:B, White:H and Red:O

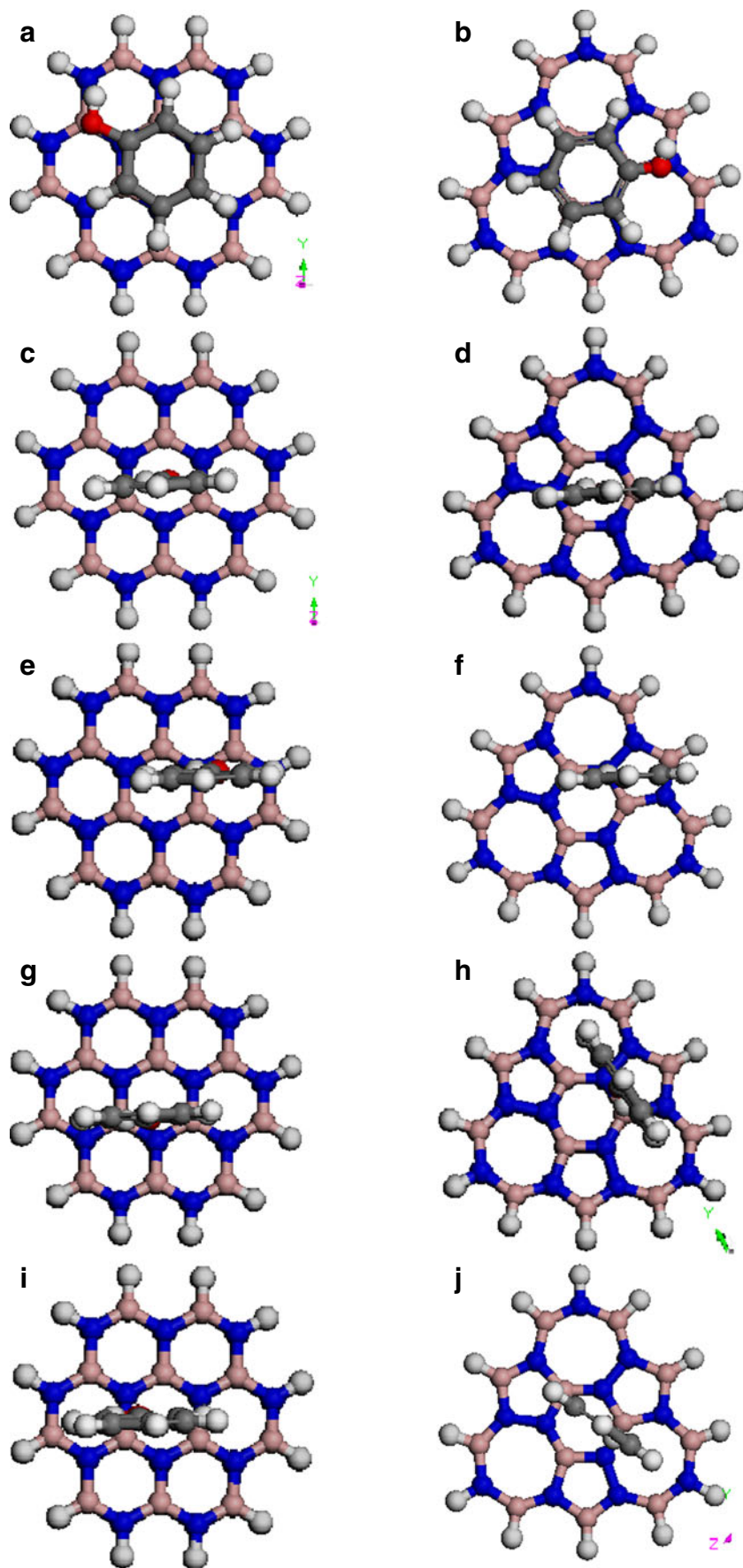
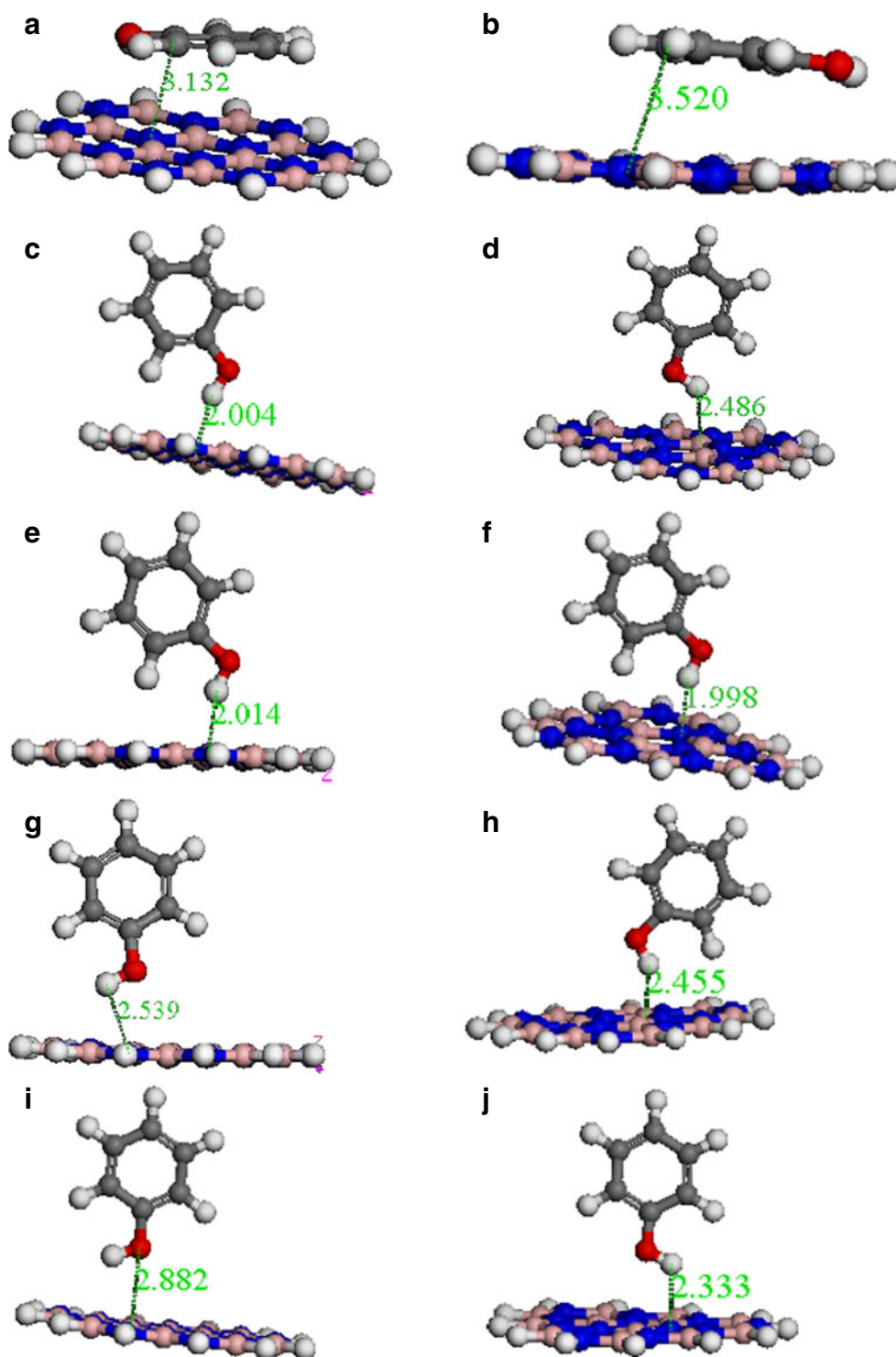


Fig. 3 These are the final geometries for (a, b) parallel cases, (c, d) perpendicular to the central hexagon, (e, f) perpendicular to the sheet at the N site, (g, h) perpendicular to the sheet at the B site, (i, j) perpendicular to the sheet at the B-N bond (bridge). Atom colors are Blue:N, Pink:B, White:H and Red:O



the molecule displays a small deviation from the planar geometry due to the OH presence. A summary of parameters is presented in Table 2.

The phenol adsorption induces an increase in the polarity of the phenol-BN structure (provided that the BN sheets are ionic systems) whose values go from 4.3×10^{-3} to $997.1 \times$

10^{-3} (neutral), to 2642.8×10^{-3} (anionic) and to 1114×10^{-3} Debye (cationic). These results also indicate that the adsorbed acid increases its charge, which is obtained from the N, which is contrary to what happens in the h-BN system. In the system at hand the dipolar moment is directed from the N-N bond (zone of higher charge) of the

Table 2 We show results of the relaxed atomic geometries, dipole moments, energy gaps and binding energies for boron nitride nanosheets

System	Bond distance (Å)					Dipolar moment (Debye) 10^{-3}	Gap (HOMO-LUMO) eV	Binding energy (eV)	Adsorption energy (eV)
	C-C	B-N	C-O	C-H	O-H				
$C_{24}H_{12}$ (Isographene) ^a	1.41 (hexagon)					1.33	0.98	8.73	
	1.39 (heptagon)								
	1.44 (pentagon)								
$C_{54}H_{18}$ [21] (Graphene)	1.41					2.9	1.94	19.18	
$C_{24}H_{12}$ [17] (Graphene)	1.41					0.2	4.06	8.99	
$B_{12}N_{12}H_{12}$ [17] (Boron nitride)		1.44				3.7	6.96	7.95	
d-BN		1.43				4.3	3.42	7.64	
[h-BN-Phenol]		1.44				1062.8	4.09	10.68	-0.47
[h-BN-Phenol] ⁻		1.44				862.3	0.05	10.68	-1.57
[h-BN-Phenol] ⁺		1.44				1921.5	4.22	10.41	-1.48
[d-BN-Phenol]		1.43				997.1	3.29	10.36	-0.63
[d-BN-Phenol] ⁻		1.43				2641.8	3.295	10.39	-1.91
[d-BN-Phenol] ⁺		1.43				1114	3.295	10.10	-2.59
Phenol	1.386		1.36	1.09	0.97	1360	4.38	2.72	
Phenol ⁺	1.41		1.31	1.09	0.98	1530	0.07	2.41	
Phenol ⁻	1.38		1.38	1.09	0.99	1230	3.51	2.67	

^a Isomer of coronene for representing graphene with lattice defects [13]

d-BN sheet to the adsorbed molecule inducing a transition from an ionic to a covalent character (as indicated by the dipole moment vector, Fig. 5a and b).

Similar to the charge effects on the atomic structure stability of the (C_6H_5OH) –h-BN, we have investigated charge effects on the (C_6H_5OH) –d-BN system. Three cases have been considered: neutral charge $[(C_6H_5OH)$ –d-BN]⁰, negative charge $[(C_6H_5OH)$ –d-BN]⁻¹ (as anionic) and positive charge $[(C_6H_5OH)$ –BN]⁺¹ (as cationic). All three cases relax to yield stable configurations. These atomic structures exhibit semiconductor behavior with energy gap of 3.29 eV (for the neutral charge) and 3.295 eV (for the anionic and cationic charge). Results of the adsorption energies are: -0.63 eV for [phenol–d-BN]⁰, -1.91 eV for [phenol–d-BN]⁻¹ and -2.59 eV for the

[phenol–d-BN]⁺¹, which indicate that phenol also is chemisorbed in all configurations like in the h-BN sheets.

Phenol adsorption in the perpendicular geometries displays interesting features, after relaxation the molecule plane is rotated with respect to the original vertical plane of the phenol: 13° for configuration 2 (Fig. 3d), 10° for configuration 3 (Fig. 3f), 5° for configuration 4 (Fig. 3h) and 7° for configuration 5 (Fig. 3j). In all cases the separation distance between the molecule and the d-BN sheet are similar (see Fig. 2). The presence of the molecule affects the d-BN sheet by inducing a small bending of the sheet, in contrast to the h-BN plane structure which remains with no changes. During the relaxation process the h-BN and d-BN sheets go through similar behavior as that shown by graphene [37] (bamboo behavior).

Fig. 4 The dipole moment vectors are for the neutral (a) phenol (the vector is on the xy plane and directed toward one of the carbons), (b) h-BN (the vector is pointing downwards to the center of the hexagon) and (c) h-BN–Phenol (the vector is pointing from the phenol toward the BN sheet). Atom colors are Grey:C, Red:O, White:H, Blue:N and Pink:B

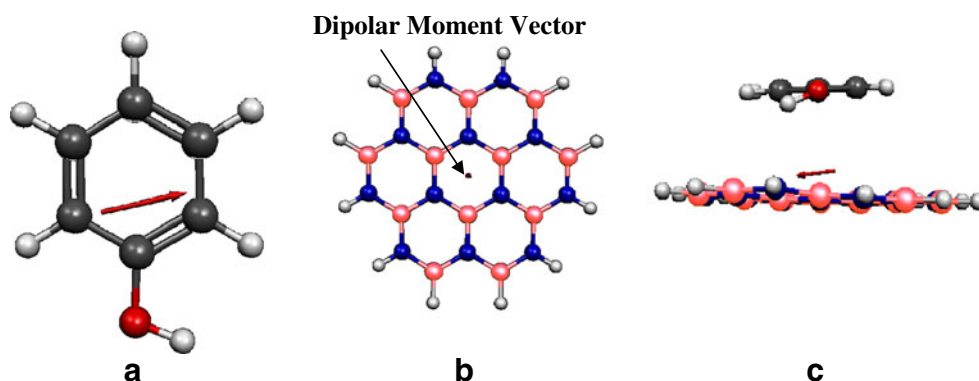
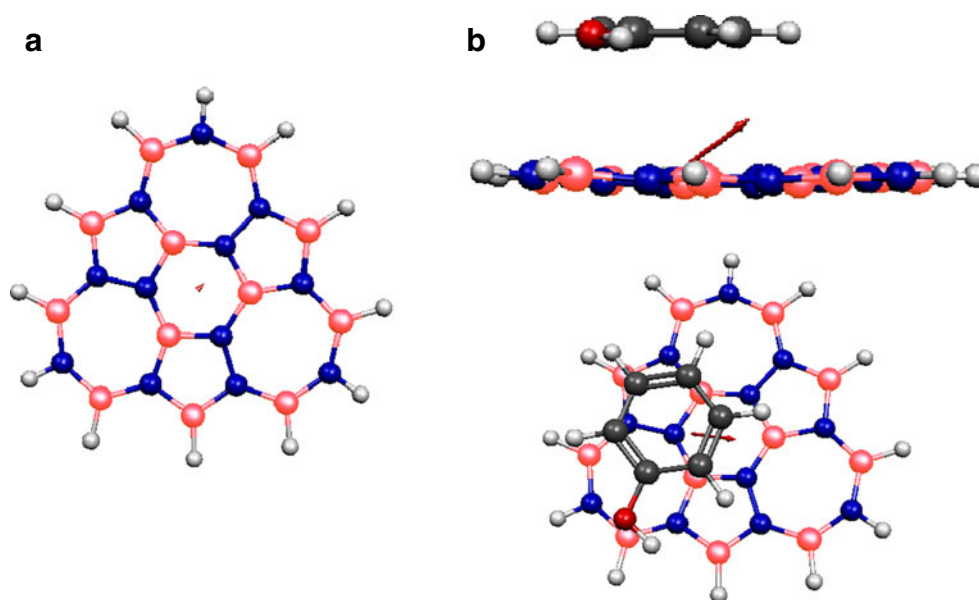


Fig. 5 Dipole moment vector for the neutral charge (a) d-BN (the vector is on the xy plane and directed toward one of the nitrogens), (b) d-BN—Phenol (the vector is directed from the hexagon of d-BN to the phenol). Atom colors are Grey:C, Red:O, White:H, Blue:N and Pink:B



An additional sixth configuration has been tested; the phenol molecule is placed within the central hexagon. This is to explore the graphene like behavior of the hexagonal h-BN sheet elasticity. During the relaxation process the B-N bond is dissociated allowing the expulsion of the phenol molecule which is accompanied with the change of the atomic orbital from sp to sp^2 . After the molecule expulsion the h-BN structure takes the original planar configuration with the phenol molecule above the central hexagon. The distance from one of the N atoms of the central hexagon to the phenol is 2.51 Å. However the total energy is larger than those obtained in all other configurations. This result is contrary to what it is obtained in the adsorption of molecules with three atoms such as O_3 . When the O_3 molecule is placed within the central hexagon, this reacts with the sheet producing a group of the epoxy type and a molecule of O_2 [3]. On the other hand when a water molecule (H_2O) is placed in a similar position, the molecule dissociates yielding a hydroxyl functional group (OH) which bonds to a boron while the other hydrogen atom forms part of the lattice [4].

Total energy results indicate that in both h-BN and d-BN sheets the phenol adsorption induces stable atomic configurations, with the parallel structure displaying the lowest minimum energy. By comparing the energy gap of the HOMO and LUMO orbitals, 6.96 eV for the phenol–h-BN structure [36] and 3.42 eV for the phenol–d-BN structure, we conclude that the structure suffers a transformation from an insulator to a semiconductor [13], as induced by the presence of defects.

The phenol adsorption on both h-BN and d-BN sheets produces an increment in the polarity of the sheets. These results indicate that the adsorbed acid increases its charge,

which is obtained from the h-BN and d-BN sheets to induce a transition from an ionic to a covalent bonding. This is similar to the phenomenon that takes place when the carboxylic acid is adsorbed on the graphene oxide in this latter case the carboxylic acid induces a larger polarity as its removal reduces the polarity by a factor of 2.33 of dipole moment [21]. In addition the adsorption energies indicate that phenol is chemisorbed in all configurations. Finally the elastic properties behave similar to that of graphene.

Conclusions

We have performed first principles total energy calculations to investigate the adsorption of phenol on the hexagonal h-BN sheets and on the modified by defects d-BN sheets. Calculations were done within the density functional theory as implemented in the DMOL code. After the atomic relaxation it is found that the minimum energy structure corresponds to that where the phenol molecule is adsorbed parallel to the BN sheets; in the hexagonal h-BN structure the phenol molecule is parallel to the central hexagon, while in the modified by defects d-BN sheets the phenol is oriented along the hexagon, a pentagon and a heptagon. The adsorption energies are -0.47 (neutral), -1.57 (anionic) and -1.48 eV (cationic) for the hexagonal h-BN configuration and -0.63 (neutral), -2.59 (anionic) and -1.91 eV (cationic) for the d-BN structure. According to these binding energies the phenol is chemisorbed, provided the adsorption energies are within the range accepted as chemisorptions. The adsorption of the molecule produces no changes in the electric conductivity in most of the cases considered here, the hexagonal h-BN and the modified by

defects d-BN structures retain their semiconductor characteristics. This is in contrast to the [phenol-h-BN]⁻¹ configuration which in turn transforms the system into a conductor. The change of polarity induces a transformation of the bonding from ionic to covalent. The elastic characteristics behave similar to that of graphene. When a phenol is placed within the central hexagon of the h-BN system during the process of relaxation the molecule is expelled, the B-N bonds are temporarily dissociated to finally form the original h-BN system with the phenol molecule adsorbed in a perpendicular direction.

Acknowledgments This work was partially supported by projects: VIEP-BUAP (CHAE-ING11-I), FIQ-BUAP (2010–2011), Cuerpo Académico Ingeniería en Materiales (BUAP-CA-177), Cuerpo Académico Física Computacional de la Materia Condensada (BUAP-CA-191) and VIEP-BUAP--EXC.

References

- Novoselov KS, Jiang D, Schedin F, Booth TJ, Khotkevich VV, Morozov SV, Geim AK (2005) *Proc Natl Acad Sci USA* 102:10451–10453
- Novoselov KS, Geim AK, Morozov SV, Jiang D, Zhang Y, Dubonos SV, Grigorieva IV, Firsov AA (2004) *Science* 306:666–669
- Chigo Anota E, Salazar Villanueva M (2009) *Sup y Vac* 22:23–28
- Chigo Anota E, Hernández Cocoltzi H, Rubio Rosas E. *Eur Phys J D* (accepted 2011)
- Young JA (2007) *J Chem Educ* 84:759
- Fraser CG (1921) *J Phys Chem* 25:1–9
- Tripathi DG (printed and published) (2008) In: Roy A (ed for and on behalf of Tulip Diagnostics(P) Ltd) *J Hyg Sci I(IV)*:1–16
- Hickey AJ, Swift D (2001) In: Baron PA Willeke K (eds) *Aerosol Measurement: Principles, Techniques, and Applications*, 2nd edn. Wiley, New York
- McDonnell G, Denver Russell AD (1999) *Clin Microbiol Rev* 12:147–179
- Kim KS, Choi SJ, Ihm SK (1983) *Ind Eng Chem Fundam* 22(2):167–172
- Hamaidi-Maouche N, Bourouina-Bacha S, Oughlis-Hammache F (2009) *J Chem Eng Data* 54:2874–2880
- Zuo X, Peng C, Huang Q, Song S, Wang L, Li D, Fan C (2009) *Nano Res* 2:617–623
- Chigo Anota E, Escobedo Morales A, Hernández Cocoltzi G. *Nanotechnol* (submitted 2011)
- Akcöltekin S, Bukowska H, Peters T, Osmani O, Monnet I, Alzahr I, d'Etat BB, Lebius H, Schleberger M (2011) *Appl Phys Lett* 98(103103):1–3
- Martínez JL, Cabria I, López MJ, Alonso JA (2009) *J Phys Chem C* 113:939–941
- You Y, Ni Z, Yu T, Shen T (2008) *App Phys Lett* 93:16311–163113
- Chigo Anota E (2009) *Sup y Vac* 22:19–23
- Nava Contreras C, Hernández Cocoltzi H, Chigo Anota E (2010) *J Mol Model*. doi:10.1007/s00894-010-0914-2
- Chigo Anota E, Salazar Villanueva M, Hernández Cocoltzi H (2010) *Phys Status Solidi C* 7:2252–2255
- Chigo Anota E, Salazar Villanueva M, Hernández Cocoltzi H (2010) *Phys Status Solidi C* 7:2559–2561
- Hernández Rosas JJ, Ramírez Gutiérrez RE, Escobedo-Morales A, Chigo Anota E (2010) *J Mol Model*. doi:10.1007/s00894-010-0818-1
- Chigo Anota E, Salazar Villanueva M, Hernández Cocoltzi H (2011) *J Nanosci Nanotechnol*. doi:10.1166/jnn.2011.3441
- Solanes Rivas CF, Hernández Cocoltzi H, Chigo Anota E. *J Mat Res* (submitted 2011)
- Solanes Rivas CF, Salazar Villanueva M, Escobedo Morales A, Chigo Anota E. *Carbon* (submitted 2011)
- Kohn W, Becke AD, Parr RG (1996) *J Phys Chem* 100:12974–12980
- Jones RO, Gunnarsson O (1989) *Rev Mod Phys* 61:689–746
- Kohn W (1999) *Rev Mod Phys* 71:1253–1266
- Chigo Anota E, Rivas Silva JF (2005) *Rev Col Fis* 37:405–419
- Delley B (1990) *J Chem Phys* 92:508–517
- Perdew JP, Wang Y (1992) *Phys Rev B* 45:13244–13249
- Delley B (1996) *J Phys Chem* 100:6107–6110
- Delley B (2000) *J Chem Phys* 113:7756–7764
- Foresman JB, Frisch Æ (1996) *Exploring chemistry with electronic structure methods*, 2nd edn. Gaussian Inc, USA, p 300
- Lin Y, Williams TV, Connell JW (2010) *J Phys Chem Lett* 1:277–283
- Zeng H, Zhi C, Zhang Z, Wei X, Wang X, Guo W, Bando Y, Golberg D (2010) *Nano Lett* 10:5049–5055
- Alem N, Erni R, Kisielowski C, Rossell MD, Gannett W, Zettl A (2009) *Phys Rev B* 80(155425):1–7
- Leenaerts O, Partoens B, Peeters FM (2008) *Appl Phys Lett* 93:193107–193109

Brønsted-Evans-Polanyi relationships for C–C bond forming and C–C bond breaking reactions in thiamine-catalyzed decarboxylation of 2-keto acids using density functional theory

Rajeev Surendran Assary · Linda J. Broadbelt ·
Larry A. Curtiss

Received: 17 February 2011 / Accepted: 22 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract The concept of generalized enzyme reactions suggests that a wide variety of substrates can undergo enzymatic transformations, including those whose biotransformation has not yet been realized. The use of quantum chemistry to evaluate kinetic feasibility is an attractive approach to identify enzymes for the proposed transformation. However, the sheer number of novel transformations that can be generated makes this impractical as a screening approach. Therefore, it is essential to develop structure/activity relationships based on quantities that are more efficient to calculate. In this work, we propose a structure/activity relationship based on the free energy of binding or reaction of non-native substrates to evaluate the catalysis relative to that of native substrates. While Brønsted-Evans-Polanyi (BEP) relationships such as that proposed here have found broad application in heterogeneous catalysis, their extension to enzymatic catalysis is limited. We report here on density functional theory (DFT) studies for C–C bond formation and C–C bond cleavage associated with the decarboxylation of six 2-keto acids by a thiamine-containing

enzyme (EC 1.2.7.1) and demonstrate a linear relationship between the free energy of reaction and the activation barrier. We then applied this relationship to predict the activation barriers of 17 chemically similar novel reactions. These calculations reveal that there is a clear correlation between the free energy of formation of the transition state and the free energy of the reaction, suggesting that this method can be further extended to predict the kinetics of novel reactions through our computational framework for discovery of novel biochemical transformations.

Keywords Enzyme catalysis · BEP relationship · Density functional theory

Introduction

Computational tools are being used increasingly to complement experimental approaches for metabolic engineering and bio-catalysis [1–10]. Recently, a computational framework based on graph theory that creates complex networks of reactions and compounds based on generalized enzyme reactions has been developed [11–18]. For a number of different systems, it has been shown that this computational framework identifies hundreds to thousands of novel biochemical pathways from a starting compound to a target compound [13, 17, 19]. Various screening methods, including pathway length, number of novel intermediates, thermodynamic landscape and pathway flux analysis [20] based on metabolic flux analysis and thermodynamic metabolic flux analysis [12], have been applied to cull this large set of novel pathways to a manageable number of attractive candidates. However, calculating the kinetic parameters for the novel reactions remains a challenge

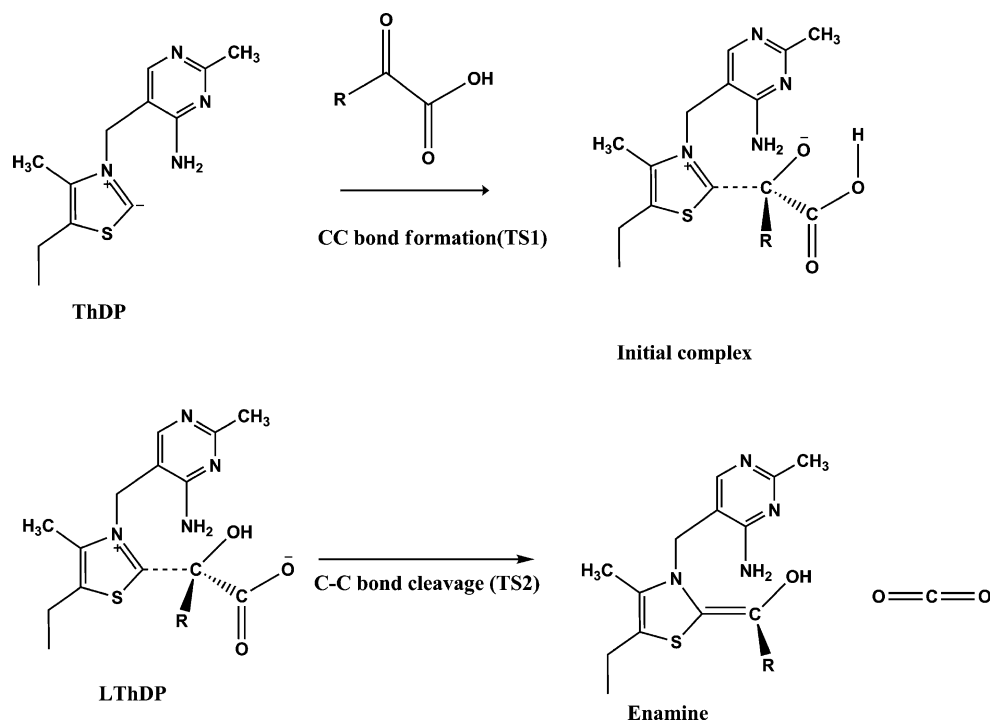
Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1062-z) contains supplementary material, which is available to authorized users.

R. S. Assary (✉) · L. J. Broadbelt
Department of Chemical and Biological Engineering,
Northwestern University,
Evanston, IL 60208, USA
e-mail: Rajeev@anl.gov

L. J. Broadbelt
e-mail: broadbelt@northwestern.edu

R. S. Assary · L. A. Curtiss
Materials Science Division and Center for Nanoscale Materials,
Argonne National Laboratory,
Argonne 60439, USA

Fig. 1 Schematic representation of the C–C bond forming and C–C bond breaking processes investigated



due to the large number of novel molecules and reactions created during network generation, the complexity of enzymatic reaction mechanisms, and the computational demands of including the active site and surroundings of the protein framework to predict the kinetics accurately. Therefore, it is essential to develop methods for predicting reaction kinetics that rely on properties that are more readily available or can be calculated with reduced demands on computational resources. This concept has found widespread use in gas phase chemistry in the form of Evans-Polanyi relationships [21–23] and in heterogeneous catalysis through Brønsted-Evans-Polanyi (BEP) relationships [24, 25].

Here, we present a computational study based on density functional theory (DFT) to calculate the kinetics and thermodynamics of the carbon–carbon bond forming and carbon–carbon bond breaking processes involved in the decarboxylation of 2-keto acids by a thiamine di-phosphate (ThDP)-containing enzyme (Fig. 1). Previously, we reported detailed energetics of the reaction mechanism, thermodynamic and kinetic feasibility of the decarboxyl-

ation of 2-keto-butanoic acid and 2-keto valeric acid compared to the native substrate, pyruvic acid, catalyzed by this enzyme (EC 1.2.7.1) [24]. In the present study, we investigated the correlation between the free energy of the reaction and the energetics of the transition state for six keto acids (pyruvic acid, 2-keto-butanoic acid, 2-keto-valeric acid, 2-keto-isovaleric acid, 2-keto-3-methyl-valeric acid and 2-keto-4-methyl-pentanoic acid), and subsequently applied this relationship to predict the activation barrier of 17 carbon–carbon bond forming and 17 carbon–carbon bond breaking reactions of similar nature. The predictions are compared to those computed directly from quantum mechanics.

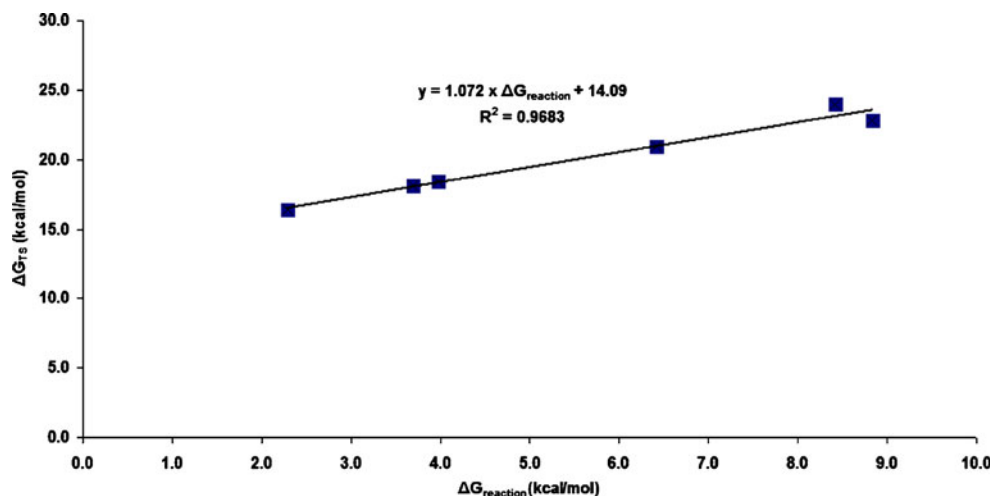
Computational details

For the quantum mechanical investigation of the mechanism of reaction, the model of the enzymatic active site shown in Fig. 1 was employed. The geometries and

Table 1 Calculated free energy of the reactions ($\Delta G_{\text{reaction}}$) and activation free energy barriers (ΔG_{TS}) for C–C bond formation and cleavage between ThDP co-factor and various 2-keto acids at 298 K in the dichloroethane dielectric using B3LYP/6-31 + G(d) level of theory. Values are reported in kcal mol^{-1}

2-Keto acid	C–C Bond formation		C–C Bond cleavage	
	$\Delta G_{\text{reaction}}$	ΔG_{TS}	$\Delta G_{\text{reaction}}$	ΔG_{TS}
Pyruvic acid	2.3	16.4	-5.2	11.0
2-Keto-butanoic acid	3.7	18.1	-7.4	9.9
2-Keto-isovaleric acid	8.8	22.8	-9.4	8.8
2-Keto-valeric acid	4.0	18.4	-6.8	10.4
2-Keto-3-methyl-valeric acid	8.4	23.9	-10.2	9.2
2-Keto-4-methyl-pentanoic acid	6.4	20.9	-9.1	9.2

Fig. 2 The Brønsted-Evans-Polanyi (BEP) relationship between the computed free energy of the reaction (ThDP–keto acid C–C bond forming) and the activation free energy barriers for the C–C bond formation for various 2-keto acids at the B3LYP/6-31 + G(d) level of theory in dichloroethane dielectric at 298 K. See Table 1 for data



energies of all species were optimized using B3LYP [26, 27] and the 6-31 + G(d) or 6-31 G(d) basis sets using Gaussian 03 [28]. Normal coordinate analysis was performed for all stationary points to characterize the transition states (one imaginary frequency) and the equilibrium structures (no imaginary frequencies) and to calculate the zero point energy (ZPE) correction and the Gibbs free energies (at 298.15 K, 1 atm). In order to account for the effects of the protein framework or the solvent, calculations were also performed using a polarizable continuum model [29–31] (C-PCM) with a dielectric constant equal to that of dichloroethane ($\epsilon = 10.36$), which is in the range reported for the protein framework [24]. Approximation of the effect of electrostatics of the protein on the electronic structure and reactivity using solvation methods is used increasingly for large proteins to avoid complicated and expensive quantum mechanics/molecular mechanics (QM/MM) calculations, and thus, this approach was adopted here [32, 33]. The structures were not re-optimized in the presence of the “solvent” since it was shown previously that re-optimization of the geometry can have a limited effect

[24]. All C-PCM calculations were carried out at 295.18 K. In order to compare the results fairly with gas phase ΔG values, a thermal correction to the free energy from the gas phase frequency results was added to the total free energy in solvent calculated with C-PCM.

Results and discussion

The computed reaction free energies and activation energy barriers for the two reaction types shown in Fig. 1 are tabulated in Table 1. For C–C bond formation (reaction 1), both the free energy of reaction (formation of C–C bond, enzyme–substrate initial complex formation) and the activation energy barrier increase as the alkyl unit adjacent to the carbonyl group of the keto acids becomes bulkier. This is due to the gradual decrease of electrophilicity of the carbonyl group due to the electron donating alkyl units, and steric interactions of the additional methyl groups adjacent to the carbonyl group at the transition state. The relationship between the free energy of the formation of the initial

Fig. 3 The BEP relationship between the computed free energy of the C–C bond breaking reactions and the activation free energy barriers for the C–C bond breaking process at the B3LYP/6-31 + G(d) level of theory in dichloroethane dielectric at 298 K

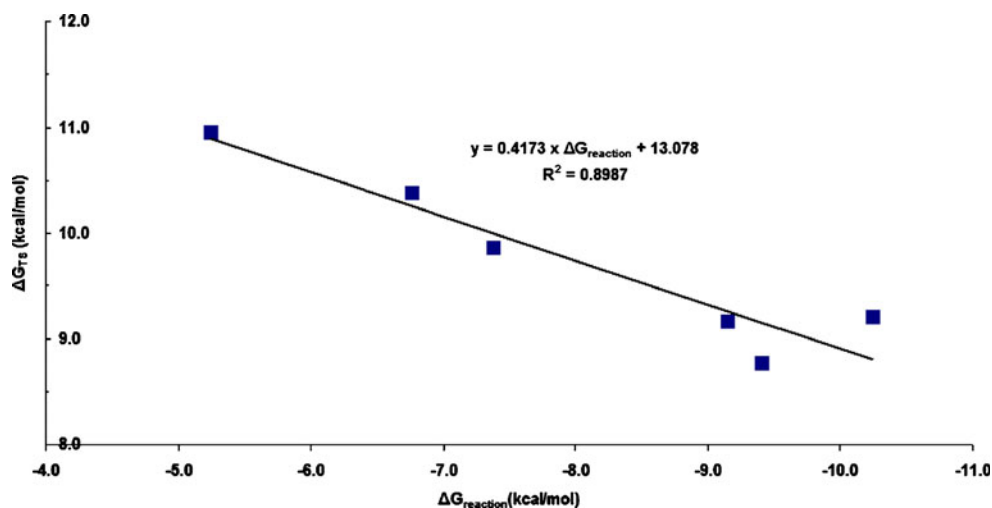


Table 2 Computed free energy of the complex formation, transition state barriers and predicted transition state barriers at the B3LYP/6-31 G(d) level of theory for the C–C bond-forming and C–C bond-breaking reactions in dichloroethane dielectric at 298 K. Chloro,

amino, methoxy and hydroxy substituted pyruvic acid (pyr), 2-keto butanoic acid (but), 2-keto-isovaleric acid (iso), and 2-keto-valeric acid (val) were used for this study. Values are reported in kcal mol⁻¹

2-Keto acid	C–C Bond formation			C–C Bond cleavage		
	$\Delta G_{\text{reaction}}$	ΔG_{TS} computed	ΔG_{TS} predicted	$\Delta G_{\text{reaction}}$	ΔG_{TS} computed	ΔG_{TS} predicted
Pyr-Cl	0.1	15.5	14.2	-11.7	4.2	8.2
Pyr-NH ₂	2.7	17.3	17.0	-4.8	9.3	11.1
Pyr-OCH ₃	-0.4	15.5	13.6	-10.6	7.4	8.6
Pyr-OH	-2.2	14.5	11.7	-10.3	7.1	8.8
But-Cl	2.1	17.0	16.3	-9.5	6.4	9.1
But-NH ₂	4.2	18.0	18.6	-8.9	7.1	9.4
But-OCH ₃	4.0	18.3	18.3	-8.8	6.8	9.4
But-OH	3.7	17.3	18.0	-9.1	6.6	9.3
Iso-Cl	5.9	22.4	20.4	-13.5	6.7	7.4
Iso-NH ₂	9.2	23.6	23.9	-9.9	7.2	8.9
Iso-OCH ₃	9.4	23.6	24.2	-11.0	6.3	8.5
Iso-OH	9.0	22.8	23.7	-10.6	7.0	8.6
Val-Cl	4.1	19.0	18.5	-9.6	6.4	9.1
Val-NH ₂	6.7	18.4	21.2	-8.3	7.2	9.6
Val-OCH ₃	4.9	19.2	19.4	-10.0	6.6	8.9
Val-OH	4.4	18.4	18.8	-8.1	7.4	9.7
Phenyl pyruvic acid	7.9	23.1	22.6	-7.4	11.1	10.0

complex between the ThDP co-factor and the keto acid, and the reaction barriers from Table 1 is shown pictorially in Fig. 2. The linear relationship observed between the free energy of the complex formation ($\Delta G_{\text{reaction}}$) and the activation energy barrier (ΔG_{TS}) in protein dielectric can be represented as

$$\Delta G_{\text{TS}} = 1.072 \times \Delta G_{\text{reaction}} + 14.09 \quad (1)$$

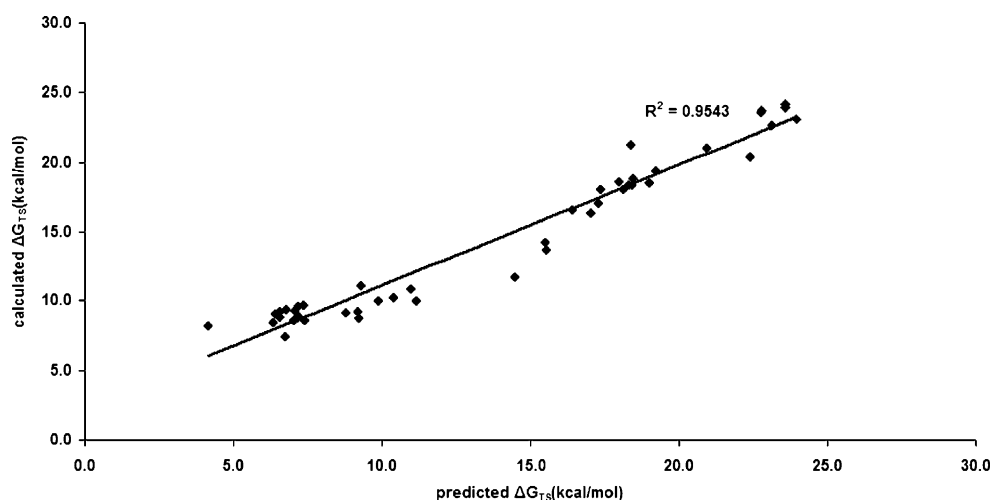
where ΔG_{TS} and $\Delta G_{\text{reaction}}$ are in kcal mol⁻¹.

Compared to the gas phase, the free energy of the complex formation reactions are marginally uphill by 1 kcal

mol⁻¹ in the presence of a protein dielectric due to the relative stabilization of the reactants compared to the product, and the activation barrier in the presence of the dielectric is approximately 4 kcal mol⁻¹ higher than in the gas phase for all keto acids (see Table S1 in the supporting information for gas phase energetics).

Similar to the C–C bond formation reaction, we derived a linear relationship between the free energy of the reaction (formation of CO₂ and enamine from the initial enzyme–substrate complex) and the activation energy barrier for the C–C bond breaking reaction (reaction 2, Fig. 1). The

Fig. 4 Comparison of calculated activation free energy barriers (B3LYP/6-31 G(d)) versus predicted activation free energy barriers for both C–C bond-formation and C–C bond breaking between 2-keto acids and ThDP co-factor in dichloroethane dielectric at 298 K



reaction barriers are proportionally lower for species with groups with higher electron donating character due to the stabilization of the positive charge at the transition state. These reactions were found to be thermodynamically downhill in both the gas phase and in the presence of the protein dielectric. The overall free energy of these reactions is less exergonic in a protein dielectric compared to the gas phase due to the added stabilization of the reactant species, and consequently the barriers are significantly larger in the protein dielectric than in the gas phase, indicating that the kinetics of these reactions are significantly controlled by the protein framework. Therefore, inclusion of accurate solvation energies is essential to derive a BEP relationship for this reaction. The derived BEP relationship for the C–C bond breaking process in protein dielectric (Fig. 3) can be expressed as

$$\Delta G_{TS} = 0.4173\Delta G_{reaction} + 13.078 \quad (2)$$

where ΔG_{TS} and $\Delta G_{reaction}$ are in kcal mol⁻¹.

In order to verify the concept of the BEP relationships, we selected 17 additional 2-keto acids and computed the free energy of the reactions and the activation barriers for both bond-forming and bond-breaking reactions at 298 K and in a protein dielectric. Compounds with chloro, amino, methoxy and hydroxyl functional groups at the terminal methyl carbon of pyruvic acid, 2-keto-butanoic acid, 2-keto-isovaleric acid, and 2-keto-valeric acid as well as phenyl pyruvic acid were selected as the substrates for the enzymatic reactions. The computed free energies ($\Delta G_{reaction}$) and activation energy barriers (ΔG_{TS}) for both C–C bond forming and C–C bond breaking reactions at 298 K in a protein dielectric for all of these substrates are tabulated in Table 2 (see Table S2 of supporting information for gas phase energetics). Also tabulated are the predicted activation free energy barriers for C–C bond formation and C–C bond breaking as calculated using the BEP relationships, Eqs. 1 and 2, respectively. The calculated activation free energy barriers and free energies of the reaction for C–C bond formation reactions for the substituted pyruvic acids were smaller than those for unsubstituted pyruvic acid, especially in the gas phase. This is generally due to the electron withdrawing effects of the chloro, methoxy and hydroxyl substituents. The computed and predicted activation free energy barriers are in reasonable agreement for the C–C bond breaking reactions (Table 2). On average, the predicted activation energy barriers are larger than the computed barriers by 2 kcal mol⁻¹ for all 17 substrates. In the gas phase, this deviation is only 1 kcal mol⁻¹ (Table S2 in Supporting Information).

Notably, the predicted activation energy barriers are in very good agreement with the computed barriers for the C–C bond formation reactions. To show this comparison more effectively, the predicted activation barriers are plotted

versus the computed activation barriers for all the 2-keto acid substrates considered in this investigation as shown in Fig. 4. The linearity of the data points with a regression coefficient of 0.95 indicates that the BEP relationship captures the C–C bond-formation well and suggests a successful interpretation of BEP relationship in addressing the C–C bond-formation reactions within the enzyme.

Summary

In summary, we have shown that BEP relationships derived from a small data set of reaction free energies and activation free energy barriers of six 2-keto acids can be applied to predict the activation energy barriers for other substrates in thiamine-containing enzymes with reasonable accuracy. This methodology can be utilized as the last step of a hierarchical screening of novel compounds and reactions to assess their kinetic feasibility given knowledge of the kinetic details of native substrates in biochemical pathways.

Acknowledgments The authors are grateful for the financial support of the National Science Foundation (CBET-0835800). This material is based upon work supported as part of the Institute for Atom-efficient Chemical Transformations (IACT), an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, and Office of Basic Energy Sciences. We gratefully acknowledge grants of computer time from the ANL Center for Nanoscale Materials.

References

1. Lee SC, Chang YJ, Shin DM, Han J, Seo MH, Fazelinia H, Maranas CD, Kim HS (2009) Designing the substrate specificity of D-hydantoinase using a rational approach. *Enzyme Microb Technol* 44:170–175
2. Fazelinia H, Cirino PC, Maranas CD (2007) Extending iterative protein redesign and optimization (IPRO) in protein library design for ligand specificity. *Biophys J* 92:2120–2130
3. Senger RS, Papoutsakis ET (2008) Genome-scale model for *Clostridium acetobutylicum*: part I. Metabolic network resolution and analysis. *Biotechnol Bioeng* 101:1036–1052
4. Senger RS, Papoutsakis ET (2008) Genome-scale model for *Clostridium acetobutylicum*: part II. Development of specific proton flux states and numerically determined sub-systems. *Biotechnol Bioeng* 101:1053–1071
5. Lee J, Yun H, Feist AM, Palsson BO, Lee SY (2008) Genome-scale reconstruction and in silico analysis of the clostridium acetobutylicum ATCC 824 metabolic network. *Appl Microbiol Biotechnol* 80:849–862
6. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA* 104:1777–1782
7. Thomas R, Paredes CJ, Mehrotra S, Hatzimanikatis V, Papoutsakis ET (2007) A model-based optimization framework for the inference of regulatory interactions using time-course DNA microarray expression data. *BMC Bioinform* 8:228–236

8. Fazelinia H, Cirino PC, Maranas CD (2009) OptGraft: a computational procedure for transferring a binding site onto an existing protein scaffold. *Prot Sci* 18:180–195
9. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox. *Nat Protoc* 2:727–738
10. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 7:129–143
11. Hatzimanikatis V, Li CH, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ (2005) Exploring the diversity of complex metabolic networks. *Bioinformatics* 21:1603–1609
12. Henry CS, Jankowski MD, Broadbelt LJ, Hatzimanikatis V (2006) Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys J* 90:1453–1461
13. Finley SD, Broadbelt LJ, Hatzimanikatis V (2009) Thermodynamic analysis of biodegradation pathways. *Biotechnol Bioeng* 103:532–541
14. Broadbelt LJ, Stark SM, Klein MT (1994) Computer generated reaction networks: on-the-fly calculation of species properties using computational quantum chemistry. *Chem Eng Sci* 49:4991–5010
15. Broadbelt LJ, Stark SM, Klein MT (1994) Computer-generated pyrolysis modeling - on-the-fly generation of species, reactions, and rates. *Ind Eng Chem Res* 33:790–799
16. Broadbelt LJ, Stark SM, Klein MT (1995) Termination of computer-generated reaction-mechanisms—species rank-based convergence criterion. *Ind Eng Chem Res* 34:2566–2573
17. Gonzalez-Lergier J, Broadbelt LJ, Hatzimanikatis V (2005) Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways. *J Chem Soc* 127:9930–9938
18. Finley SD, Broadbelt LJ, Hatzimanikatis V (2009) Thermodynamic analysis of biodegradation pathways. *Biotechnol Bioeng* 104:1086–1097
19. Henry CS, Broadbelt LJ, Hatzimanikatis V (2010) Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. *Biotechnol Bioeng* 106:462–473
20. Wiback SJ, Famili I, Greenberg HJ, Palsson BØ (2004) Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J Theor Biol* 228:437–447
21. Broadbelt LJ, Klein MT, Dean BD, Andrews SM (1995) Structure/reactivity relationships for high-performance polyamides—kinetics of the reactions of N,N'-dihexylphthalamides in the presence of added copper iodide and water. *J Polym Sci A Polym Chem* 33:533–545
22. Evans MG, Polanyi M (1938) Inertia and driving force of chemical reactions. *Trans Faraday Soc* 34:11–24
23. Osuna S, Houk KN (2009) Cycloaddition reactions of butadiene and 1,3-dipoles to curved arenes, fullerenes, and nanotubes: theoretical evaluation of the role of distortion energies on activation barriers. *Chem Eur J* 15:13219–13231
24. Assary RS, Broadbelt LA (2010) Computational screening of novel thiamine-catalyzed decarboxylation reaction of 2-keto acids. *Bioprocess Biosyst Eng* 34:375–388
25. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic-behavior. *Phys Rev Abstr* 38:3098–3100
26. Becke AD (1993) Density-functional thermochemistry.3. The role of exact exchange. *J Chem Phys* 98:5648–5652
27. Lee CT, Yang WT, Parr RG (1988) Development of the collesalvetti correlation-energy formula into a functional of the electron-density. *Phys Rev B* 37:785–789
28. Frisch MJ et al (2003) Gaussian03, revision E.01. Gaussian Inc, Wallingford
29. Takano Y, Houk KN (2005) Benchmarking the conductor-like polarizable continuum model (CPCM) for aqueous solvation free energies of neutral and ionic organic molecules. *J Chem Theor Comput* 1:70–77
30. Cossi M, Rega N, Scalmani G, Barone V (2003) Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J Comput Chem* 24:669–681
31. Barone V, Cossi M (1998) Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J Phys Chem A* 102:1995–2001
32. Wang J, Dong H, Li S, He H (2005) Theoretical study toward understanding the catalytic mechanism of pyruvate decarboxylase. *J Phys Chem B* 109:18664–18672
33. Siegbahn P, Himo F (2009) Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *J Biol Inorg Chem* 14:643–651

Virtual screening and in vitro assay of potential drug like inhibitors from spices against glutathione-S-transferase of filarial nematodes

Shamina Azeez · Rosana O. Babu · Riju Aykkal ·
Reena Narayanan

Received: 3 May 2010 / Accepted: 9 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract Glutathione-S-transferase(s) (GST) enzyme from *Brugia malayi* has been exploited as a target in lymphatic filariasis therapeutics. An active GST is a homodimer of a 208 residue long monomer consisting of two domains, a smaller α/β domain and a larger α domain. The components of the glutathione (GSH) system, mainly GST enzymes, are critical antioxidant and detoxification system responsible for the long-term existence of filarial worms in mammalian host; hence they are major chemotherapeutic targets in filarial species. In the present study, 58 phytochemicals from 10 plants, predicted and reported to have potential nematicidal activity and ADMET satisfaction, have been docked to GST enzyme of *B. malayi* to assess their binding affinity and consequently their inhibitory activity. A comparative study has been made with commonly employed chemotherapeutic GST inhibitors such as cibacron-blue, butylated hydroxyanisole, hexyl glutathione and ethacrynic acid. In vitro effects of potential drug like compound from *in silico* results have been done for validation of docking studies. In vitro assay revealed efficacy in GST inhibition in the following compounds: linalool (97.50%), *alpha*-pinene (90.00%), strychnine (87.49%), vanillin (84.99%), piperine (79.99%), isoeugenol (62.49%), curcumin (57.49%), *beta*-caryophyllene (39.50%), cinnamic acid (27.49%), capsaicin (19.99%), citronellol (19.99%) and geraniol (17.49%). An online database (www.spicebioinfo.res.in/gstleadbase) has been developed, which will serve as a useful repository of information on GST inhibitors for future development of drugs against filarial nematodes. These findings thus

suggest that the above phytochemicals could be potentially developed as lead molecules for targeting GST of lymphatic filarial parasites.

Keywords *Brugia malayi* · *Dirofilaria immitis* · Docking · Glutathione S-transferase(s) · Phytochemicals

Introduction

Lymphatic filariasis (LF) is a mosquito-borne tropical disease caused by the nematode parasites *Wuchereria bancrofti*, *Brugia malayi* and *B. timori* [1]. It is the major cause of acute and chronic morbidity in 81 countries in Asia-Pacific, Africa and the Americas. Approximately 1.3 billion people living in these regions are at risk of infection [2]. The adult parasites live 5 to 10 years, of which the fecund life span is 4 to 6 years. Several hundreds to thousands of infective mosquito bites are necessary to establish infection. Of these, three parasites *W. bancrofti* accounts for nearly 90% of LF infections worldwide. *B. malayi* is prevalent only in some parts of South and Southeast Asia, and *B. timori* is found only in Indonesia. The drugs used for treating LF include annual doses of diethylcarbamazine (DEC), DEC plus albendazole, or ivermectin plus albendazole; none of these is effective in killing adult worms, and treatments are therefore aimed at reducing transmission and pathology [3, 4]. Since an effective treatment for filarial adult worms is currently unavailable, new chemical classes of compounds with macrofilaricidal activities are now required [5]. Recently Srinivasan et al. [6] reported ethacrynic acid, plumbagin and curcumin as inhibitory compounds against GSTs of bovine filarial worms *Setaria digitata*. Only a few studies have reported the use of phytochemicals as GST inhibitors.

S. Azeez (✉) · R. O. Babu · R. Aykkal · R. Narayanan
Indian Institute of Spices Research,
Calicut 673012, Kerala, India
e-mail: shamina@spices.res.in

The components of the glutathione (GSH) system GSTs (glutathione-S-transferase(s)) and GSHPx (glutathione peroxidases) are the major defense systems present in filarial nematodes. The role of this secreted enzyme is the inhibition of the oxidative burst of leukocytes and neutralization of secondary products of lipid peroxidation, thus providing an explanation for the resistance of these parasites to immune effector mechanisms and their persistence in the mammalian host [7]. The mechanism of action of GST(s) (E.C.2.5.1.18), a large family of multifunctional dimeric enzymes includes defense against oxidative attack via conjugation of electrophiles to glutathione and reduction of lipid hydroperoxides [8]. Due to their primary role in drug metabolism, GSTs have been the recent focus of research as a potential drug target for anti-schistosomal [9], antimalarial [10, 11], and antifilarial [12–14] drug development. In addition to their isomerization and GSH conjugation activities, in mammals these enzymes contribute to defense against oxidative stress, by virtue of both their selenium-independent GSH peroxidase activities [15]. Inhibition of parasitic GST affects the survival of the parasites or helps in the enhancement of activity of presently available antifilarial drugs [16]. GST from human filarial parasites is significantly different from human GST in sequence and structure [17]. Hence *B. malayi* GST was exploited to design new target based chemotherapeutic agents.

An active GST is a homodimer of a 208 residue long monomer consisting of two domains (smaller α/β domain and larger α domain) (Fig. 1). The N-terminal small domain (residues 1 to 74) is an α/β structure with the folding topology $\beta\alpha\beta\alpha\beta\alpha$ arranged in the order β_2 , β_1 , β_3 and β_4 with β_3 anti-parallel to the others, forming a regular β -sheet with a right-handed twist surrounded by three α -helices. The C terminal, large domain 2 (82–208 residues) is α -helical. The residues that interface the two $\beta\alpha\beta$ and $\beta\beta\alpha$ motifs are Trp38, Phe8, Val33, Cys47, Leu52 and Leu43 in human π GST. In Bm-GST the residues Val33, Cys47 and Leu43 are replaced by Ile 38, Phe47 and Met43 [17]. The secondary structure of *B. malayi* GST has been generated by GenTHREADER – Protein fold recognition software (<http://bioinf.cs.ucl.ac.uk/threader/>) [18].

A wide range of chemical compounds including alkaloids, coumarins, flavonoids, benzofurans, terpenoids and steroids have been isolated from various plant extracts and these have been found to possess various pharmacological, nematocidal and insecticidal activities. A comprehensive review of chemical constituents and pharmacological profiles of 10 selected medicinal plants, including spices, have led to the identification of potential nematocides. These new nematocides of natural origin may lead to higher safety and efficiency in nematode control and nematocidal drug development.

Historically, herbs, shrubs and spices have enjoyed a rich tradition of use for their flavor enhancement characteristics and medicinal properties [19, 20]. Spices hold the promise of providing both significant clinical benefits and key insights into the pathophysiology of cancer, arthritis, inflammation, respiratory disorders, gastrointestinal disturbances, allergy and microbial infections. Numerous demonstrations of preclinical efficacy of turmeric and various spices in animal models for preventing cancer and cardiovascular disorders have been reported [21]. This study was conducted with the objective of exploring the nematocidal activity of herbs and spices, with special reference to its potential to inhibit GST activity.

Materials and methods

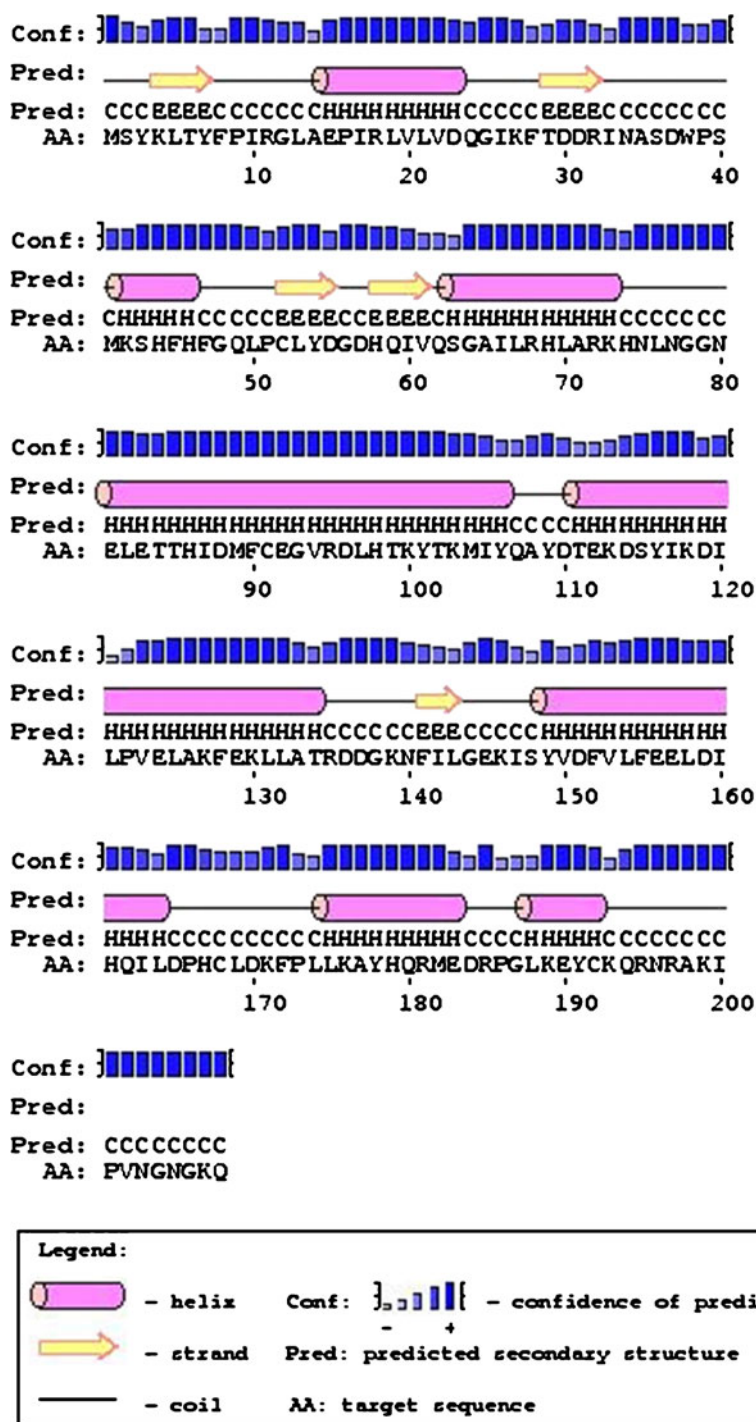
Database screening and activity prediction

Initially, a review of plants with nematocidal property was made, which helped us identify 10 spices and medicinal plants: coriander, cassia, turmeric, allspice, cinnamon, strychnous, lemongrass, garlic, litsea and vanilla. The chemical compounds from these plants were collected through literature search and from Dr. Duke's phytochemical and ethno-botanical databases (<http://ars-grin.gov/duke/>). The screening results revealed the presence of 128 nematocidal phytochemicals in these plants. The PASS server [22] was used to predict nematocidal activity and GST substrate activity of the phytochemicals (<http://195.178.207.233/PASS/AP.html>); PreADMET server (<http://preadmet.bmcrcd.org/>) was used to predict the drug-likeness and ADME-Tox (Absorption, Distribution, Metabolism Excretion and Toxicity) properties [23]. The ADME-Tox properties of a compound together with its pharmacological properties such as drug likeness are conventionally a part of drug development. The compounds obeying the ADMET rules and drug likeness rules were short listed for docking studies.

Ligand structure

The canonical smiles notations of phytochemicals were collected from PubChem (<http://pubchem.ncbi.nlm.nih.gov/>), ChemSpider (<http://chemspider.com>) and DrugBank (<http://www.drugbank.ca/>). The 3D structures of compounds were developed by 3D Structure Generator CORINA [24, 25] using canonical smiles of the compound. Energy minimization and molecular optimization of all compounds were done using Arguslab 4.0.1 [26]. Geometry optimization was carried out using AM1 (Austin Model 1), semi-empirical quantum mechanics force field in Arguslab4.0.1. The best conformer thus obtained was based on energy minimization

Fig. 1 Secondary structure of *Brugia malayi* GST: Generated by GenTHREADER – Protein fold recognition software (<http://bioinf.cs.ucl.ac.uk/threader/>) [18]



and geometry optimization. The final structures exhibiting lowest energy were saved in *.pdb format for input in to MVD environment.

Target protein structure

Theoretically solved structure of *Brugia malayi* glutathione-S-transferase was selected as the target for docking study, since to date there are no experimentally

solved structures for *Wuchereria bancrofti*, *Brugia malayi* or *B. timori*. The GST structure was downloaded from Protein Data Bank (PDB id - 1SJO) and the structure optimized using Swiss PDB viewer software. Three active sites were detected in the GST enzyme of *B. malayi* using Molegro Virtual Docker. Active site residues of the GST enzyme were predicted using WHATIF server (<http://swift.cmbi.ru.nl/servers/html/index.html>) [27]. Active site residues predicted by WHATIF were Tyr7, Tyr101, Tyr106,

Table 1 Characteristics of phytochemicals docked with GST of *B. malayi*

Serial number	Ligand	Moldock score (kJ/mol)	H-bond interaction energy (kJ/mol)	Number of H- bonds
1.	1,8-Cineole	-51.9902	-0.948	1
2.	2-Furfuraldehyde	-64.8983	-4.449	3
3.	2-Methoxycinnamaldehyde	-64.8564	-2.953	3
4.	Acetyl-eugenol	-82.2233	-2.196	4
5.	Alpha-copaene	-53.9326	0	0
6.	Alpha-humulene	-68.188	0	0
7.	Alpha-pinene	-75.5588	-2.772	3
8.	Alpha-terpinene	-58.5699	0	0
9.	Alpha-terpineol	-59.8115	-2.688	2
10.	Benzaldehyde	-58.3635	-2.5	2
11.	Brucine	-91.7289	-4.112	6
12.	Brucine-n-oxide	-95.1105	-4.2539	6
13.	Capsaicin	-58.5662	-2.4545	3
14.	Carvacrol	-63.5588	-2.772	3
15.	Cinnamaldehyde	-66.6676	-2.5	2
16.	Cinnamic acid	-66.1044	-1.657	1
17.	Cinnamyl acetate	-71.9436	-1.588	3
18.	Cinnamyl alcohol	-62.8854	-2.326	2
19.	Cis asarone	-66.448	-1.465	3
20.	Citral	-67.7443	-2.141	2
21.	Citronellal	-70.7249	-2.231	1
22.	Citronellol	-73.133	-2.5	2
23.	Curcumin	-137.66	-8.4343	7
24.	Decanal	-71.6631	-2.5	1
25.	Diabolone	-67.3305	-0.783	2
26.	Diallyl disulfide	-59.5614	0	0
27.	Diallyl trisulfide	-61.3798	0	0
28.	Diallylsulfide	-56.062	0	0
29.	Dodecanal	-81.4478	-2.5	1
30.	Eugenol	-76.089	-5.745	3
31.	Genostychnine	-86.3784	-2.000	4
32.	Geraniol	-69.8422	-2.5	2
33.	Icajine	-83.174	-3.109	4
34.	Isoeugenol	-75.9879	-5.557	3
35.	Isopulegone	-57.0261	-1.219	1
36.	Limonene	-58.8357	0	0
37.	Linalool	-80.895	-2.240	2
38.	Methyl-eugenol	-70.2178	-1.423	2
39.	Methyl-isoeugenol	-66.8614	-1.522	3
40.	Myristicin	-69.8614	-1.7384	2
41.	Neral	-72.4322	0	1
42.	Nonanal	-65.4356	0	1
43.	NVA	-78.8236	-5.557	3
44.	Octanal	-62.3393	0	0
45.	p-cymene	-58.6187	0	0
46.	Piperine	-79.985	-3.139	4
47.	pseudo strychnine	-78.8246	-1.350	4
48.	Strychnine	-84.0994	-1.436	3
49.	trans- 2-decan-1-ol	-75.983	-2.5	2

Table 1 (continued)

Serial number	Ligand	Moldock score (kJ/mol)	H-bond interaction energy (kJ/mol)	Number of H- bonds
50.	trans-anethole	-64.2715	-1.221	2
51.	Turmerone	-79.985	-1.653	2
52.	Undecanal	-79.5591	-2.5	1
53.	Vanillin	-89.5321	-7.0792	5
54.	Verbenol	-53.0257	-2.357	2
55.	Vomicine	-79.4487	-5.469	6
56.	Zingiberene	-74.8799	0	0
57.	β -caryophyllene	-51.5322	0	0
58.	β -colubrine	-91.7316	-1.247	4

Phe8, Phe45, Phe47, Phe155, Pro9, Pro51, Pro201, Ile10, Ile33, Ile105, Ile200, Arg11, Arg32, Arg95, Gly12, Gly48, Gly64, Leu13, Leu50, Asn34, Asn203, Ala35, Try38, Lys42, Lys103, Gln49, Gln62, Ser63, His98, Thr99, Thr102, Asp159, Val61, and Val202. Docking was carried out using single active-site having large volume (79.608 Å³) among the three cavities. This cavity was chosen since it binds glutathione with the highest specificity compared to the other sites, when docked with the whole protein using MVD. Amino acid residues present in the active site selected for docking and grid generation were Gln49, Gln12, Ile10, Ile105, Gly12, Ser63, His98, Tyr7, Tyr101, Tyr106, Pro9, Pro51, Pro201, Arg11, Arg32, Arg95, Thr99, Thr102, Val61, Val202, Lys42 and Lys103.

Molecular docking

Molecular docking study was carried out by using Molegro Virtual Docker [28]. The entire protein structure was loaded on to MVD platform for docking process. MVD performs flexible ligand docking, so the optimal geometry of the ligand is determined during the docking. MVD includes MolDock Score [28] and PLANTS Score [29] for evaluating docking solutions. MVD returns multiple poses representing different potential binding modes. This can be useful when the best-scoring (i.e., lowest-energy) pose does not represent the native binding mode or when multiple binding modes exist. Here clustering has been

used to reduce the number of poses found during the docking run and only the most promising ones are reported. Compounds with the lowest dock score and high interaction with active-site was taken for in vitro studies based on the availability of the compound.

In vitro GST assay

The phytochemicals β -caryophyllene, capsaicin, cinnamic acid, citronellol, curcumin, eugenol, geraniol, isoeugenol, linalool, myristicin, neral, α -pinene, piperine, terpineol, vanillin and strychnine were purchased in the pure form from Sigma Chemicals, USA; glutathione (GSH) and 1-chloro-2, 4-dinitrobenzene (CDNB) were purchased from Sisco Research Laboratories Pvt. Ltd., (Mumbai, India). *Dirofilaria immitis* microfilaria, the canine filarial nematode used for in vitro study, was obtained from the District Veterinary Centre Campus, Calicut, Kerala.

GST crude enzyme was obtained by centrifuging the serum containing ~2000 filarial nematodes at 1000 rpm for 2 min, and washing twice with phosphate buffered saline (PBS) at pH 7.4. The nematodes were ground with micro pestle and glass powder. The solution was centrifuged at 10000 rpm at 4 °C for 30 min. Supernatant was dialyzed against PBS overnight and made up to 2 ml. The following phytochemicals were used to study their GST inhibitory activity, at a concentration of 0.001 mg ml⁻¹ in ethanol: β -caryophyllene, capsaicin, cinnamic acid, citronellol, curcumin, eugenol, geraniol,

Table 2 Binding energy scores of GST-inhibitors

Serial number	GST-inhibitor	Dock score (kJ/mol)	H-bond interaction energy (kJ/mol)	Number of H-bonds
1.	Cibacron-blue	-129.656	-1.561	5
2.	Hexyl glutathione	-113.777	-6.320	10
3.	Ethacrynic acid	-87.569	-2.424	5
4.	Butylated hydroxyanisole	-68.431	-3.640	4

isoeugenol, linalool, myristicin, neral, α -pinene, piperine, terpineol, vanillin and strychnine (dissolved in water).

The dialyzed enzyme fraction (0.1 ml) was incubated in the presence of 1 ml of 0.001 mg ml⁻¹ concentration of the phytochemicals listed above, in the presence of 1 mM

glutathione reduced (GSH), and 0.1 M phosphate buffer, pH 6.5, for 1 hour at room temperature. A control containing ethanol was also maintained. GST activity was measured using the method of Habig et al. [30], by initiating the reaction with the addition of 1 mM 1-chloro-2,4-dinitroben-

Fig. 2 Docking view showing Hydrogen bond interaction of ligands with residues in active site of GST enzyme. **(a)** curcumin, **(b)** brucine-n-oxide, **(c)** beta-colubrine, **(d)** brucine, **(e)** genostychnine, **(f)** strychnine, **(g)** vanillin and **(h)** linalool

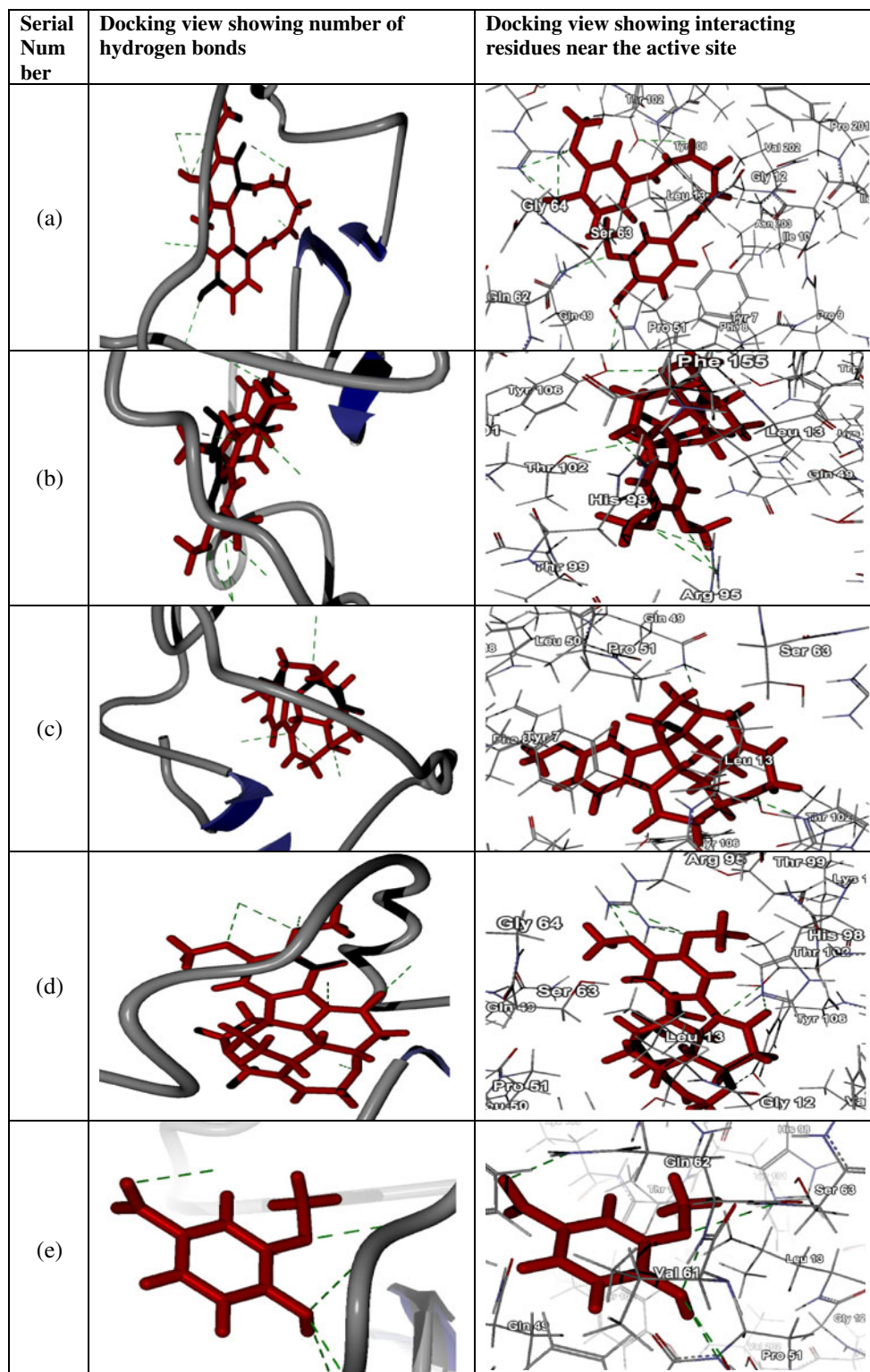
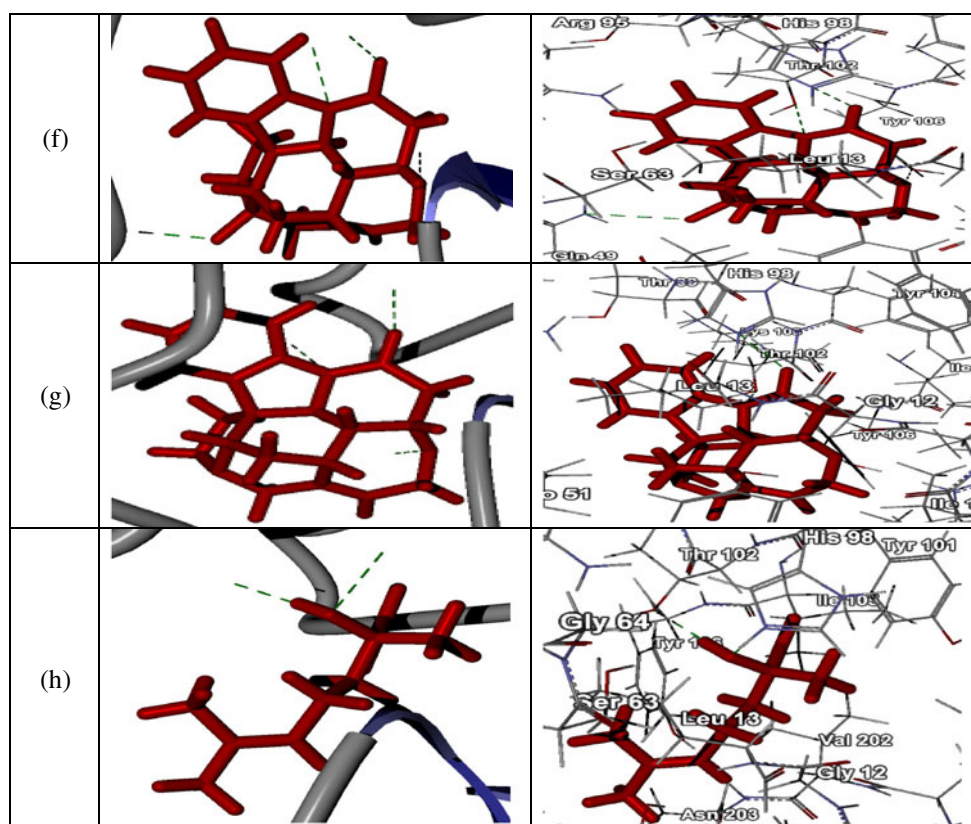


Fig. 2 (continued)



zene (CDNB) and following the change in absorbance at 340 nm, in a Shimadzu 1601 UV-Visible spectrophotometer. The GST activity was expressed as change in absorbance at 340 nm per minute per ml crude enzyme extract. Two replicates of each treatment were maintained.

Results and discussion

Biological activity prediction yielded 58 of the 128 phytochemicals with nematocidal, anti-helminthic and GST

substrate activities. These phytochemicals also satisfied both ADME-Tox and drug likeness rules and were selected for docking studies. Docking results showed that all 58 compounds docked satisfactorily to the GST enzyme active site with good docking scores of less than -51.532 kcal mol⁻¹. Hence these phytochemicals of comparatively less docking energy and greater number of hydrogen bond interactions were selected as promising lead compounds after docking studies (Table 1).

An *in silico* study was performed to compare the binding affinity of commonly employed chemotherapeutic GST

Table 3 Binding energy scores of eight phytochemicals and interacting residues

Serial number	Phytochemical	Dock score (kJ/mol)	H bond interaction energy (kJ/mol)	Common interacting residues	Number of H bonds
1.	Curcumin	-137.66	-8.4343	Gln62, Gln64, Pro51, Val202, Tyr106, Pro201, Tyr7	7
2.	Brucine N oxide	-95.1105	-4.2539	Thr102, His98, Arg95 (3 H bonds), Tyr106	6
3.	beta-Colubriline	-91.7316	-1.247	Tyr7, Gln49, Thr102, Tyr106	4
4.	Brucine	-91.7289	-4.112	Arg95 (2 H bonds), Thr102 (2 H bonds), Tyr106, His98	6
5.	Vanillin	-89.5321	-7.0792	Gln62 (2 H bonds), Ser63, Pro51(2 H bonds)	5
6.	Genostrychnine	-86.3784	-2.000	Gln49, His98, Thr102, Tyr106	4
7.	Strychnine	-84.0994	-1.436	Thr102, Tyr106, His98	3
8.	Linalool	-80.895	-2.240	His98, Thr102	2


```

BmGST_Y12788      MSYKLTYPPIRGLAEPiRLVLVDQGIKFTDDRINASDWPSPKSHFHFQGLPCLYDGDHQI 60
WbGST_AY195867   MSYKLTYPPIRGLAEPiRLVLVDQGIKFTDDRINASDWPSPKSHFHFQGLPCLYDGDHQI 60
DiGST_P46426     MSYKLTYPPIRGLAEPiRLLLVDQGIKFTDEHLPKDDFVSIKSKQFQGLPCLYDGDQOI 60
                  *****:*****:* .*: *:*:*:*****:*****:
BmGST_Y12788      VQSGAILRHLARKHNLNGGNELETHIDMFCEGVRDLHTKYTKMIYQAYDTEKDSYIKDI 120
WbGST_AY195867   VQSGAILRHLARKHNLNGGNELETHIDMFCEGIRDLHTKYAKMIYQAYDTEKDSYIKDI 120
DiGST_P46426     VQSGAILRHLARKFNLNGENNAETSYVDMFYEGIRDLHSKYTRMIYQAYETQKDFPIKNI 120
                  *****:*****:*****:*****:*****:*****:*****:
BmGST_Y12788      LPVEAAKFEKFLATRDDGKNFILGEEKISYVDFVLFEELDIHQILDPHCLDKFPLLKAYHQ 180
WbGST_AY195867   LPVELAKFEKLLATRDDGKNFILGEEKISYVDFVLFEELDIHQILDPHCLDKFPLLKAYHQ 180
DiGST_P46426     LPQELAKLEKLLATRDNNGKNFILGDKISFADYVLFEELDVQQILDPHCLEKFPPLLKAFHQ 180
                  ** * *:***:*****:*****:*****:*.:*****:*****:*****:***
BmGST_Y12788      RMEDRPGLEKEYCKQRNRAKIPVNGNGKQ 208
WbGST_AY195867   RMEDRPGLEKEYCKQRNRAKIPVNGNGKQ 208
DiGST_P46426     RLGDKPKIKEYCAKRNASKMPVNGNGKQ 208
                  *: *:* :***** :* *:*:*****

```

Fig. 3 Multiple sequence alignment (ClustalW) of GST sequences of *B. malayi*, *D. immitis* and *W. bancrofti*, which show that these sequences are identical. BmGST (Y12788) and WbGST (AY195867) show 98% similarity, BmGST and DiGST (P46426) share 74%

similarity and WbGST and DiGST 75% similarity. The (*) denotes identical bases, (:) denotes strongly similar amino acids and (.) denotes weakly similar amino acids

inhibitor substances such as cibacron-blue, butylated hydroxyanisole, hexyl glutathione and ethacrynic acid, with the phytochemicals used for this study. Several potent phytochemicals that possess docking scores very similar to the current GST inhibitor drugs were identified (Table 2). Among the GST inhibitors butylated hydroxyanisole (BHA) markedly reduces worm viability [31]. BHA had a docking score of $-68.431 \text{ kcal mol}^{-1}$;

phytochemicals with lower docking score than BHA are potential GST inhibitors, and represent promising starting points as lead compounds to treat LF. Thus phytochemicals which exhibit low dock scores and strong hydrogen bond interaction energy and greater number of hydrogen bonds in docking studies such as curcumin (PubChem CID: 969516; MW: $368.380 \text{ g mol}^{-1}$), vanillin (PubChem CID: 1183; MW: $152.147 \text{ g mol}^{-1}$), strychnine (PubChem CID: 5979; MW: $334.412 \text{ g mol}^{-1}$), genostrychnine (PubChem CID: 73393; MW: $350.411 \text{ g mol}^{-1}$), brucine (PubChem CID: 442021; MW: $394.464 \text{ g mol}^{-1}$), brucine-n-oxide (PubChem CID: 161215; MW: $410.463 \text{ g mol}^{-1}$), beta-columbrine (PubChem CID: 10512; MW: $364.438 \text{ g mol}^{-1}$) and linalool (PubChem CID: 6549; MW: $154.249 \text{ g mol}^{-1}$) are promising hits as GST inhibitors of natural origin. The hydrogen bond interaction of these lead compounds with the target residues is shown in Fig. 2. The analysis showed that curcumin has greater number of H-bond interactions and strychnine the least. Table 3 shows interacting properties of the eight highly docked phytochemicals to the target protein.

Table 4 GST activity in *Dirofilaria immitis*, treated with phytochemicals

Serial number	Treatments	GST-activity (units ^a)	Inhibition (% of control)
1.	Control	13.333	
2.	Linalool	0.333	97.50
3.	Alpha-pinene	1.333	90.00
4.	Strychnine	1.667	87.49
5.	Vanillin	2	84.99
6.	Piperine	2.667	79.99
7.	Isoeugenol	5	62.49
8.	Curcumin	5.667	57.49
9.	Beta-Caryophyllene	8.333	37.50
10.	Cinnamic acid	9.667	27.49
11.	Capsaicin	10.667	19.99
12.	Citronellol	10.667	19.99
13.	Geraniol	11	17.49
14.	Alpha-Terpineol	ND	
15.	Neral	ND	
16.	Myristicin	ND	

^a Units = $\times 10^{-3}$ dA/minute/ml crude enzyme extract

ND=Not Detectable

GST is extensively investigated as a major target against several parasitic infections [31–35]. GST of filarial nematodes has very similar function and multiple sequence analysis revealed its similarity in sequence level (Fig. 3). There is no sequence in public domain of GST of *B. timori* to compare its relatedness to other LF GSTs. GST protein of *B. malayi* (BmGST) and *W. bancrofti* (WbGST) shares 98% similarity and GST of the canine filariasis nematode, *Dirofilaria immitis* (DiGST) shares only 74% similarity to BmGST and 75% to WbGST. We used *D. immitis* GST for in vitro studies, since the other filarial nematodes were not available.

In vitro studies indicated that linalool (97.50%), *alpha*-pinene (90.00%), strychnine (87.49%), vanillin (84.99%), piperine (79.99%), isoeugenol (62.49%), curcumin (57.49%), *beta*-caryophyllene (39.50%), cinnamic acid (27.49%), capsaicin (19.99%), citronellol (19.99%) and geraniol (17.49%) have good potential as nematicidal compounds against filarial GST (Table 4). Molecular structures of these compounds are given in Fig. 4.

Terpineol, neral and myristicin had no detectable inhibitory effect. These in vitro studies help validate the results obtained from *in silico* docking studies. The reason why in vitro activities do not correlate closely with *in silico* docking could be because we have used *B. malayi* GST for *in silico* studies, while in vitro studies were carried out with *D. immitis*, due to the difficulty in obtaining samples of *B. malayi*. Multiple sequence alignment of GST

Fig. 4 Molecular structures of assayed compounds - Structures drawn using ChemSketch software

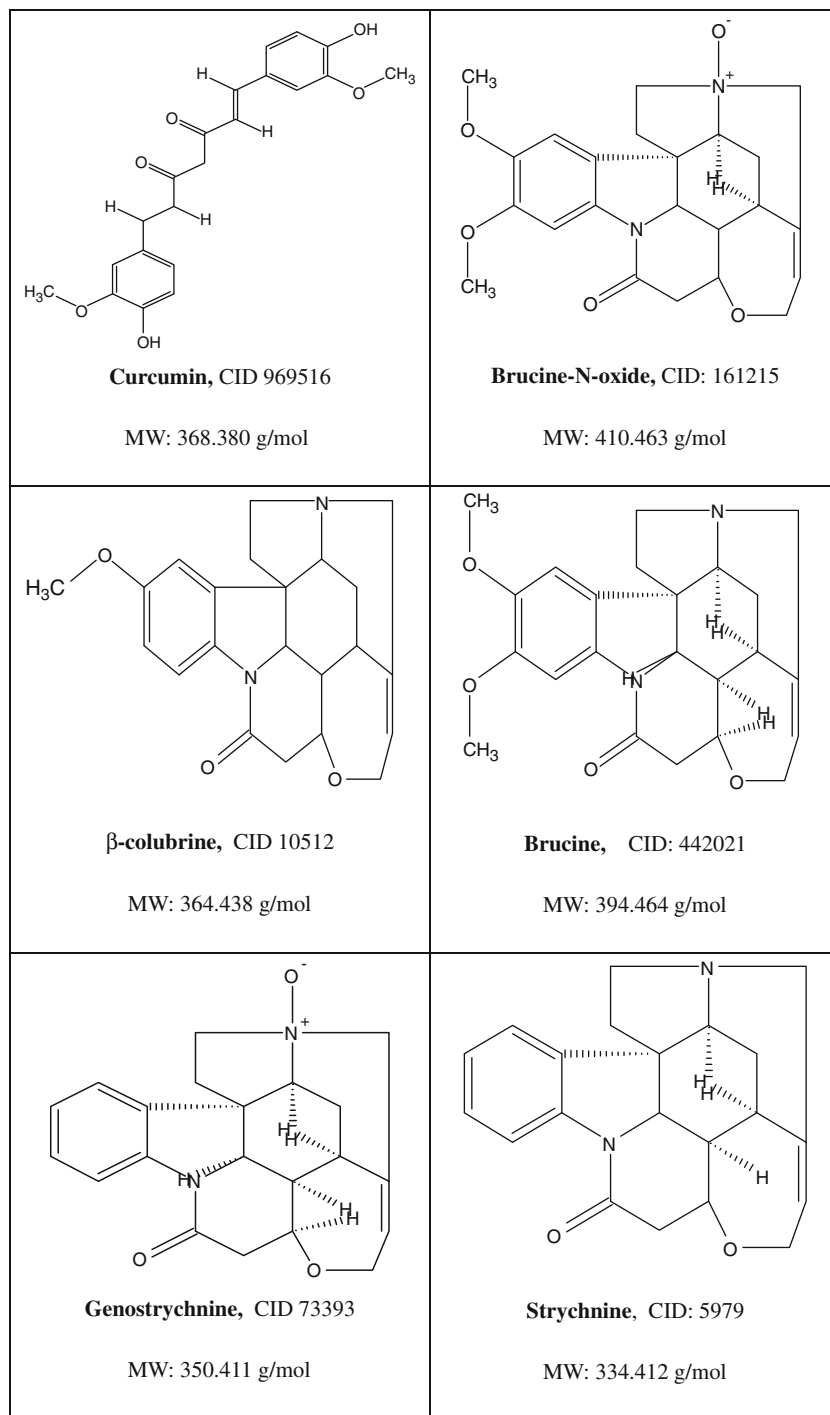
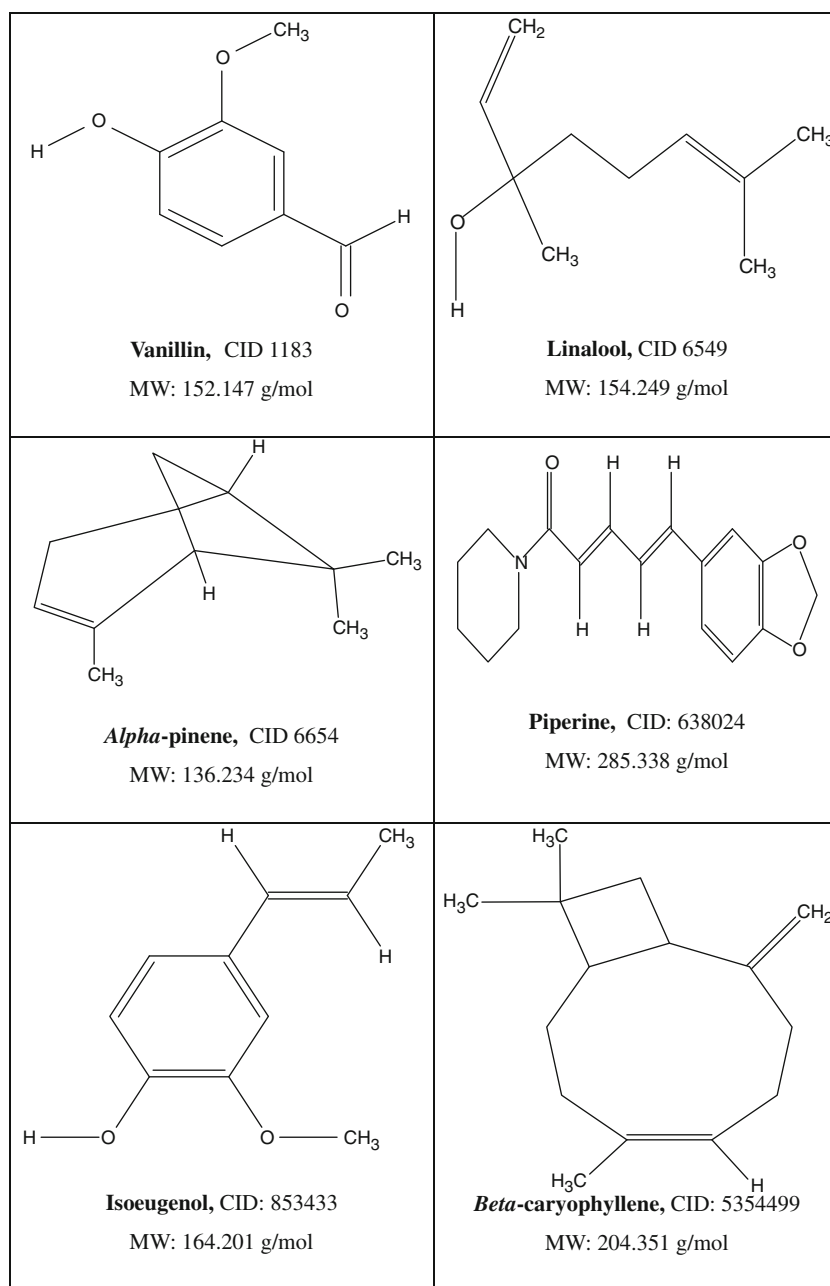


Fig. 4 (continued)

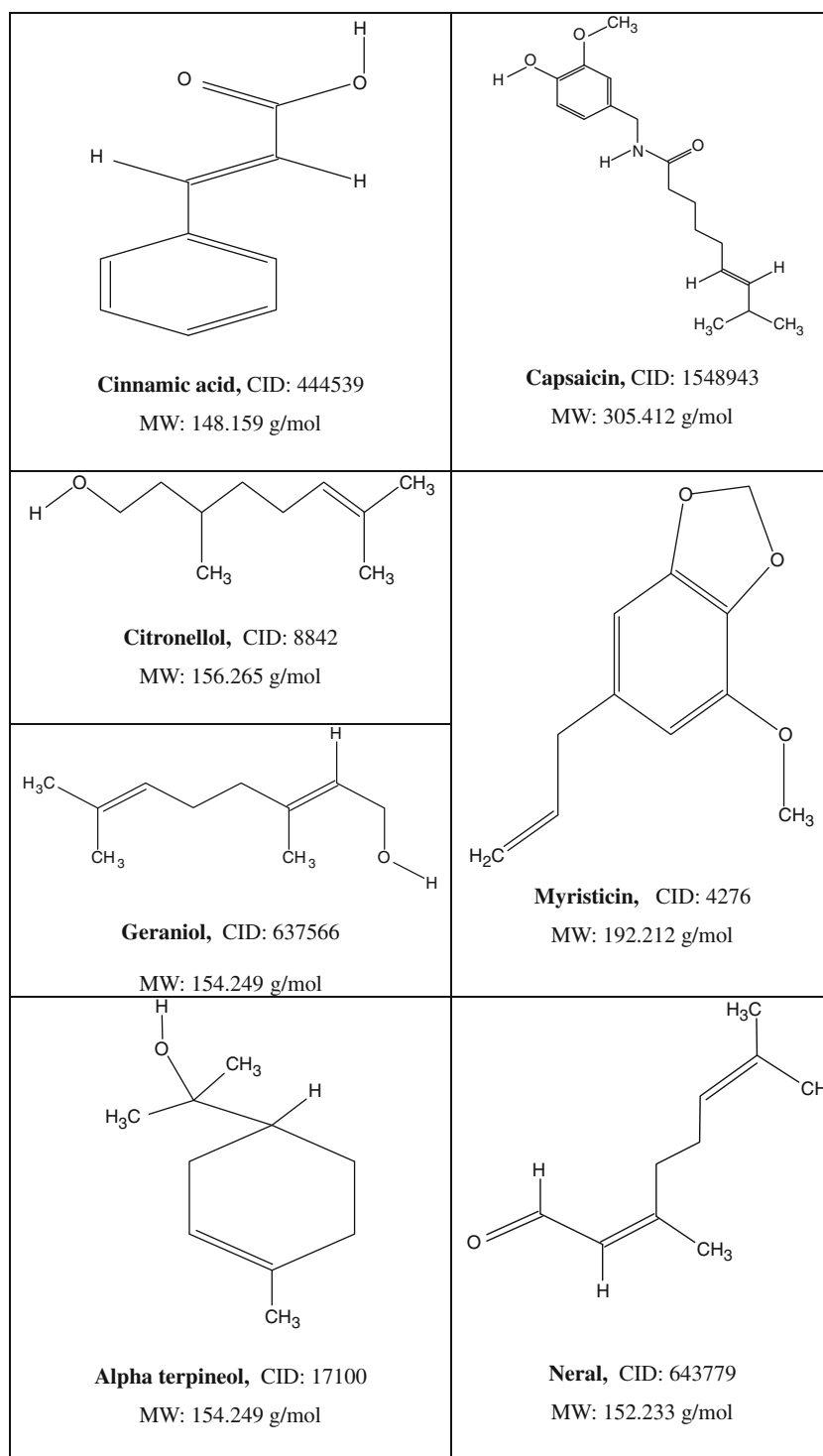


sequences revealed that BmGST and DiGST (P46426) share only 74% similarity (Fig. 3). The difference in the correlation between the *in silico* and *in vitro* results could be attributed to the structural differences among the BmGST and DiGST.

Curcumin is being used for treatment of cancer, wounds and as a cosmetic [36] among other medicinal uses. GST inhibiting activity of curcumin has been identified in various organisms [37, 38] and its worm motility inhibition was found to be effective at 54.29 μM [6]; the dried seed of *Strychnos nux-vomica* L., has been effectively used in Chinese folk medicine for the treatment of liver cancer and associated pathological abnormalities for ages [39]. Vanilla is

a valued spice for its aroma and flavor. The anti-inflammatory activity of linalool has been reported earlier [40]. Piperine, a bioavailability enhancer from black pepper (*Piper* spp.), has already been reported to inhibit glucuronidation activity in rats and guinea pigs [41]. Singh et al. [42] reported that piperine inhibited rat hepatocyte-mediated glucuronidation of 3-hydroxybenzo[a]pyrene with an IC_{50} of 50 $\mu\text{mol L}^{-1}$. Co-administration of piperine and curcumin to humans and rats enhanced the bioavailability of curcumin by 2000% and 154%, respectively [43]. Isoeugenol is a well-known antioxidant and its other biological activities include anti-inflammatory, antibiotic, antioxidant, anti-carcinogenic and local anaesthetic activities.

Fig. 4 (continued)



Since the above studied compounds are of natural origin, which satisfies both ADMET and drug likeness properties, these compounds can be used as potent lead compounds against filarial parasites. An online database (www.spicebioinfo.res.in/gstleadbase) has been developed, which it is hoped will serve as a useful repository of information on GST inhibitors for future development of drugs against filarial nematodes.

Conclusions

To summarize, we have employed virtual screening protocol, molecular docking to identify potential drug-like inhibitors of the detoxifying enzyme - GST - of *Brugia malayi*. Several potential drug-like inhibitors have been screened and found to interact with GST satisfactorily. Phytochemicals like curcumin, brucine-n-oxide, *beta*-colu-

brine, brucine, genostrychnine, strychnine, vanillin and linalool revealed strong binding with less docking scores and more number of hydrogen bond interactions to GST of *B. malayi*. This *in vitro* and *in silico* docking study validates GST inhibitory activity of compounds such as linalool, *alpha*-pinene, strychnine, vanillin, piperine, isoeugenol, curcumin, *beta*-caryophyllene, cinnamic acid, capsaicin, citronellol and geraniol, hence these compounds are novel, alternative drug therapy, of natural origin, for treatment of filariasis through inhibition of GST. Further studies are required to mark them as lead compounds for the development of novel drugs against lymphatic filariasis.

Acknowledgments We acknowledge with thanks the facilities provided by Director, Indian Institute of Spices Research (IISR), Calicut; Dr. Santhosh J. Eapen, Co-ordinator, Distributed Information Sub-Centre (DBT, New Delhi), of IISR (Indian Institute of Spices Research), Dr. O.K. Sindu, District Veterinary Centre Campus, Calicut, who supplied the filarial nematodes *Dirofilaria immitis* and the assistance rendered by Ms. M. Dinsha, Senior Research Fellow, IISR.

References

- WHO (2002) Lymphatic Filariasis: The Disease and Its Control. Technical Report 71. Geneva
- WHO (2008) Global programme to eliminate lymphatic filariasis. Wkly Epidemiol Rec 83:333–341
- Liu LX, Weller PF (1996) Antiparasitic drugs. N Engl J Med 334:1178–1184
- Oliveira-Menezes A, Lins R, Norões J, Dreyer G, Lanfredi RM (2007) Comparative analysis of a chemotherapy effect on the cuticular surface of *Wuchereria bancrofti* adult worms *in vivo*. Parasitol Res 101:1311–1317
- Bajpai P, Verma SK, Katiyar D, Tewari N, Tripathi RP, Bansal I, Saxena JK, Misra-Bhattacharya S (2005) Search for new prototypes for the chemotherapy of filariasis: a chemotherapeutic and biochemical approach. Parasitol Res 95:383–390
- Srinivasan L, Mathew N, Muthuswamy K (2009) *In vitro* antifilarial activity of glutathione-S-transferase inhibitors. Parasitol Res 105:1179–1182
- Cookson E, Blaxter ML, Selkirk ME (1992) Identification of the major soluble cuticular glycoprotein of lymphatic filarial nematode parasites (gp29) as a secretory homolog of glutathione peroxidase (filariasis/*Brugia*/antioxidant/surface protein). Proc Natl Acad Sci USA 89:5837–5841
- Ahmad R, Srivastava AK (2008) Inhibition of glutathione-S-transferase from *Plasmodium yoelii* by protoporphyrin IX, cibacron blue and menadione: implications and therapeutic benefits. Parasitol Res 102:805–807
- Michele AM, DeWight RW, John AT (1995) Crystal structures of aschistosomal drug and vaccine target: glutathione S-transferase from *Schistosoma japonica* and its complex with the leading antischistosomal drug Praziquantel. J Mol Biol 246:21–27
- Harwaldt P, Rahlfs S, Becker K (2002) Glutathione S-transferase of the malarial parasite *Plasmodium falciparum*: characterization of a potential drug target. Biol Chem 383:821–830
- Ahmad R, Srivastava AK (2007) Purification and biochemical characterization of cytosolic glutathione-S-transferase from malarial parasites *Plasmodium yoelii*. Parasitol Res 100:581–588
- Brophy PM, Campbell AM, van Eldik AJ, Teesdale-Spittle PH, Liebau E, Wang MF (2000) Beta-carbonyl substituted glutathione conjugates as inhibitors of *O. volvulus* GST2. Bioorg Med Chem Lett 10:979–981
- Liebau E, Wildenburg G, Brophy PM, Walter RD, Henkle-Duhrsen K (1996) Biochemical analysis, gene structure and localization of the 24 kDa glutathione S-transferase from *Onchocerca volvulus*. Mol Biochem Parasitol 80:27–39
- Rao UR, Salinas G, Mehta K, Klei TR (2000) Identification and localization of glutathione S-transferase as a potential target enzyme in *Brugia* species. Parasitol Res 86:908–915
- Zhao TJ, Singhal SS, Piper TJ, Cheng JZ, Pandya U, Clark-wronski J, Awasthi S, Awasthi YC (1999) The role of human glutathione-S-transferases hGSTA1-1 and hGSTA2-2 in protection against oxidative stress. Arch Biochem Biophys 367:216–224
- Lüersen K, Walter RD, Müller S (1998) The putative γ -glutamylcysteine synthetase from *Plasmodium falciparum* contains large insertions and a variable tandem repeat. Mol Biochem Parasitol 98:131–142
- Bhargavi R, Vishwakarma S, Murty US (2005) Modeling analysis of GST (glutathione-S-transferases) from *Wuchereria bancrofti* and *Brugia malayi*. Bioinform 1:25–27
- Jones DT (1999) GenTHREADER: an efficient and reliable protein folds recognition method for genomic sequences. J Mol Biol 287:797–815
- Okunade AL (2002) *Ageratum conyzoides* L. (Asteraceae). Fitoterapia 73:1–16
- Kaefer CM, Milner JA (2008) The role of herbs and spices in cancer prevention. J Nutr Biochem 19:347–361
- Luthra PM, Singh R, Chandra R (2001) Therapeutic uses of *Curcuma longa* (Turmeric). Indian J Clin Biochem 16:153–160
- Filimonov DA, Poroikov VV (1996) PASS: Computerized prediction of biological activity spectra for chemical substances. Bioactive Compound Design. Possibilities for Industrial Use. BIOS Scientific, Oxford, pp 47–56
- Lee Sung Kwang (2005) *In silico* high-throughput screening for ADME/Tox properties: PreADMET program. Abstr Conf Comb Chem Jpn 21:22–28
- Schönberger H, Schwab CH, Hirsch AJ, Gasteiger J (2000) Molecular modelling of fullerene dendrimers. J Mol Model 6:379–395
- Sadowski J, Gasteiger J, Klebe G (1994) Comparison of automatic three-dimensional model builders using 639 X-ray structures. J Chem Inf Comput Sci 34:1000–1008
- Peng C, Ayali PY, Schlegel HB, Frisch MJ (1995) Using redundant internal coordinates to optimize equilibrium geometries and transition states. J Comput Chem 16:49–51
- Hendlich M, Rippmann F, Barnickel G (1997) Mapping of protein surface cavities and prediction of enzyme class by a self-organizing neural network. J Mol Graph 15(359–363):389
- Thomsen R, Christensen MH (2006) MolDock: a new technique for high-accuracy molecular docking. J Med Chem 49:3315–3321
- Korb O, Stutzle T, Exner TE (2009) Empirical scoring functions for advanced protein-ligand docking with PLANTS. J Chem Inf Model 49:84–96
- Habig WH, Pabst MJ, Jakoby WB (1974) Glutathione S-transferases. The first enzymatic step in mercapturic acid formation. J Biol Chem 249:7130–7139
- Gupta S, Bhandari YP, Reddy MV, Harinath BC, Rathaur S (2005) *Setaria cervi*: immunoprophylactic potential of glutathione-S-transferase against filarial parasite *Brugia malayi*. Exp Parasitol 109:252–255
- Morrison CA, Colin T, Sexton JL, Bowen F, Wicker J, Friedel T, Spithill TW (1996) Protection of cattle against *Fasciola hepatica* infection by vaccination with glutathione S-transferase. Vaccine 14:1603–1612

33. Grezel D, Capron M, Grzych JM, Fontaine J, Lecocq JP, Capron A (1993) Protective immunity induced in rat schistosomiasis by a single dose of the Sm28 GST recombinant antigen: effector mechanisms involving IgE and IgA antibodies. *Eur J Immunol* 23:454–460
34. Sexton JL, Milner AR, Panaccio M, Waddington J, Wijffels G, Chandler D, Thompson C, Wilson L, Spithill TW, Mitchell GF, Campbell NJ (1990) Glutathione S-transferase: novel vaccine against *Fasciola hepatica* infection in sheep. *J Immunol* 145:3905–3910
35. Veerapathran A, Dakshinamoorthy G, Gnanasekar M, Reddy MVR, Kalyanasundaram R (2009) Evaluation of *Wuchereria bancrofti* GST as a vaccine candidate for lymphatic filariasis. *PLoS Negl Trop Dis* 3:e457
36. Kawamori T, Lubet R, Steele VE, Kelloff GJ, Kaskey RB, Rao CV, Reddy BS (1999) Chemopreventive effect of curcumin, a naturally occurring anti-inflammatory agent, during the promotion/progression stages of colon cancer. *Cancer Res* 59:597–601
37. Oetari S, Sudibyo M, Commandeur JNM, Samhoedi R, Vermeulen NPE (1996) Effects of curcumin on cytochrome P450 and glutathione S-transferase activities in rat liver. *Biochem Pharmacol* 51:39–45
38. Awasthi S, Pandya U, Singhal SS, Lin JT, Thivyanathan V, Seifert WE Jr, Awasthi YC, Ansari GAS (2000) Curcumin–glutathione interactions and the role of human glutathione S-transferase P1-1. *Chem Biol Interact* 128:19–38
39. Deng XK, Yin W, Li WD, Yin FZ, Lu XY, Zhang XC, Hua ZC, Cai BC (2006) The anti-tumor effects of alkaloids from the seeds of *Strychnos nux-vomica* on HepG2 cells and its possible mechanism. *J Ethnopharmacol* 106:179–186
40. Peana AT, D'aquila PS, Panin F, Serra G, Pippia P, Moretti MDL (2002) Antiinflammatory activity of linalool and linalyl acetate constituents of essential oils. *Phytomedicine* 9:721–726
41. Reen RK, Jamwal DS, Taneja SC, Koul JL, Dubey RK, Wiebel FJ, Singh J (1993) Impairment of UDP-glucose dehydrogenase and glucuronidation activities in liver and small intestine of rat and guinea pig in vitro by piperine. *Biochem Pharmacol* 46:229–238
42. Singh J, Dubey RK, Atal CK (1986) Piperine-mediated inhibition of glucuronidation activity in isolated epithelial cells of the guinea-pig small intestine: evidence that piperine lowers the endogenous UDP-glucuronic acid content. *J Pharmacol Exp Ther* 236:488–489
43. Shoba G, Joy D, Joseph T, Majeed M, Rajendran R, Srinivas PS (1998) Influence of piperine on the pharmacokinetics of curcumin in animals and human volunteers. *Planta Med* 64:353–356

Theoretical investigations of a high density cage compound 10-(1-nitro-1, 2, 3, 4-tetraazol-5-yl) methyl-2, 4, 6, 8, 12-hexanitrohexaazaisowurtzitane

Jian-ying Zhang · Hong-chen Du · Fang Wang ·
Xue-dong Gong · Yin-sheng Huang

Received: 16 January 2011 / Accepted: 22 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract A new polynitro cage compound with the framework of HNIW and a tetrazole unit, i.e., 10-(1-nitro-1, 2, 3, 4-tetraazol-5-yl) methyl-2, 4, 6, 8, 12-hexanitrohexaazaisowurtzitane (NTz-HNIW) has been proposed and studied by density functional theory (DFT) and molecular mechanics methods. Properties such as IR spectrum, heat of formation, thermodynamic properties, and crystal structure were predicted. The compound belongs to the *Pbca* space group, with the lattice parameters $a=15.07$ Å, $b=12.56$ Å, $c=18.34$ Å, $Z=8$, and $\rho=1.990$ g·cm⁻³. The stability of the compound was evaluated by the bond dissociation energies and results showed that the first step of pyrolysis is the rupture of the N–NO₂ bond in the side chain. The detonation properties were estimated by the Kamlet-Jacobs equations based on the calculated crystal density and heat of formation, and the results were 9.240 km·s⁻¹ for detonation velocity and 40.136 GPa for detonation pressure. The designed compound has high thermal stability and good detonation properties and is probably a promising high energy density compound (HEDC).

Keywords Crystal structure · Detonation performance · DFT · Molecular mechanics · Stability

Introduction

Nowadays, searching for novel high energy density materials (HEDMs) to meet the future energetic and defense demands has become one of the most active regions and seems to be more and more urgent. To find new energetic compounds with superior performances, one promising approach is to explore compounds with compact structure and high nitrogen content, which has been proved by many recent researches and development efforts [1–4]. Highly nitrated cage molecules such as hexanitrohexaazaisowurtzitane (HNIW) [5], polynitroadamantanes (PNAs) [6, 7], octanitrocubane (ONC) [8–10], and 4-trinitroethyl-2, 6, 8, 10, 12-pentanitrohexaazaisowurtzitane (TNE-HNIW) [11, 12] have therefore gained much attention in recent years. These strained cage compounds possess a concomitant increase in heat of formation (HOF) and density, which make them powerful explosives [13–17].

Experiment showed that TNE-HNIW, a derivative of HNIW, successfully prepared recently, has higher stability than HNIW [11] and is a potential explosive. Theoretical studies not only support the experimental result, but also show that it has better detonation properties than HNIW [12]. Considering the comprehensive advantages of tetrazole, such as high nitrogen content, high enthalpy of formation, and high thermal stability, a new derivative of HNIW with a tetrazole unit, 10-(1-nitro-1, 2, 3, 4-tetraazol-5-yl) methyl-2, 4, 6, 8, 12-hexanitrohexaazaisowurtzitane (NTz-HNIW) was proposed (Fig. 1). We expect the structure will possess similar or even better properties than HNIW. With the help of quantum chemistry calculations,

J.-y. Zhang (✉) · H.-c. Du · F. Wang · X.-d. Gong · Y.-s. Huang
School of Chemical Engineering,
Nanjing University of Science and Technology,
Nanjing 210094, People's Republic of China
e-mail: nanjinger@163.com

X.-d. Gong
e-mail: gongxd325@mail.njust.edu.cn

Y.-s. Huang
e-mail: huangyinsheng@sina.com

X.-d. Gong
State Key Laboratory of Explosion Science and Technology,
Beijing Institute of Technology,
Beijing 100081, People's Republic of China

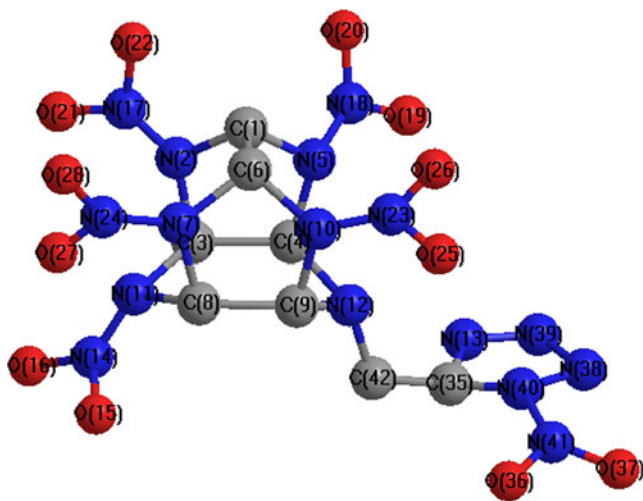


Fig. 1 Molecular structure of NTz-HNIW (hydrogen atoms omitted)

the structural and energetic properties, such as HOF, thermodynamic properties, crystal density, detonation performance, and thermal stability are evaluated theoretically, which help us to understand the designed structure deeply and to find its possibility as a candidate of HEDC.

Computational methods and details

The stable geometry, vibrational spectrum, and related properties were investigated using the Gaussian 03 [18] and Materials Studio 4.4 packages [19]. The B3LYP [20, 21] method of density functional theory (DFT) with the standard Gaussian basis set 6-31 G(d,p) [22] which has been proved credible for evaluating the structure and energetic properties of various chemical systems [23–25] has been adopted to optimize the molecular structure. Vibrational analysis and natural bond orbital (NBO) analysis [26] have been carried out too at the same level.

Detonation velocity and detonation pressure, the important parameters reflecting the explosive performance of energetic materials, were predicted using the empirical Kamlet-Jacobs equations [27] as follows:

$$D = 1.01 \left(N \bar{M}^{-1/2} Q^{1/2} \right)^{1/2} (1 + 1.30\rho), \quad (1)$$

$$P = 1.558\rho^2 N \bar{M}^{1/2} Q^{1/2}, \quad (2)$$

where D is detonation velocity ($\text{km}\cdot\text{s}^{-1}$), P is detonation pressure (GPa), ρ is the density of explosive ($\text{g}\cdot\text{cm}^{-3}$), N is the moles of gaseous detonation products per gram of explosives, \bar{M} is the average molecular weight of the

detonation products, and Q is the detonation energy ($\text{cal}\cdot\text{g}^{-1}$). N , \bar{M} , and Q are determined based on the most exothermic principle.

Pyrolysis mechanism and thermal stability were evaluated using bond dissociation energy (BDE), the difference between the energies of the parent molecule and the corresponding radicals in the unimolecular bond dissociation [28–30]. In the present paper, BDE of the initial bonds were calculated using the following equation:

$$BDE(R_1R_2) = [\Delta_f H(R_1) + \Delta_f H(R_2)] - \Delta_f H(R_1R_2), \quad (3)$$

where R_1 and R_2 are the radicals produced by the breaking of R_1 – R_2 bond of R_1R_2 compound [31].

The possible polymorph and crystal structure parameters of the title compound were predicted by searching the possible packings among seven most probable space groups ($P2_1/c$, $P-1$, $P2_12_12_1$, $Pbca$, $C2/c$, $P2_1$, and $Pna2_1$) [32–36] using molecular mechanics method.

Results and discussion

Infrared spectrum

IR spectrum is not only the basic property of compounds, but also an effective measure to identify substances. Besides, it has a direct relation with and is necessary to predict the thermodynamic properties. Therefore, it is of great significance to calculate the IR spectrum for the title compound. Figure 2 presents the simulated IR spectrum based on the scaled harmonic vibrational frequencies obtained at the B3LYP/6-31 G (d, p) level which has been

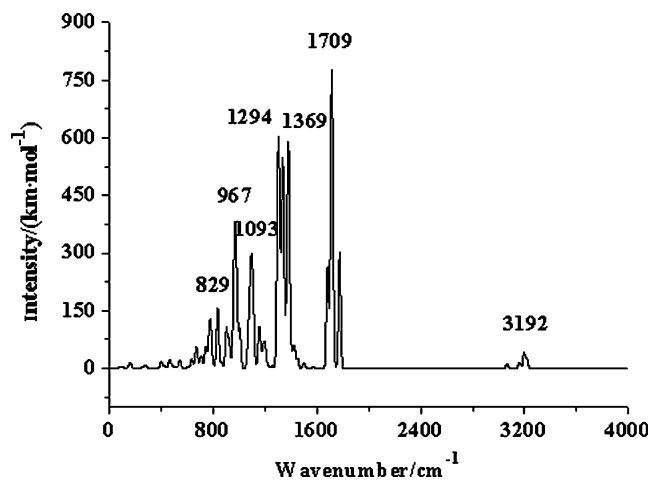


Fig. 2 The simulated infrared spectrum for NTz-HNIW

Table 1 Calculated energies of species

Compd.	C(g)	H ₂	O ₂	N ₂	NTz-HNIW	TNE-HNIW	HNIW
E_0^a /(a.u.)	-37.84628	-1.16837	-150.31626	-109.51853	-2087.25893	-2278.50069	-1791.18314
HOF/(kJ·mol ⁻¹)	710.52 ^b	0	0	0	755.21	705.61 ^c	691.30 ^d

^a E_0 is the total energy after the correction of zero-point energy

^b from Ref. [44]

^c from Ref.[12]

^d from Ref.[45]

proved to be suitable for the computation of IR spectrum [37–40]. Obviously, there are seven main characteristic regions. The modes in 3041~3236 cm⁻¹ are associated with the C–H stretch. In this region the strongest characteristic peak is at 3192 cm⁻¹. The remarkable signal at 1709 cm⁻¹ is associated with the N=O asymmetric stretch of nitro groups and the strong characteristic peak at 1369 cm⁻¹ is characterized to the N=N symmetric stretch of the tetrazole ring and C–H wagging in plane. Band at 1294 cm⁻¹ is a complex of C–H twisting out of plane and wagging in plane. The peak at 1093 cm⁻¹ is composed of the N–N asymmetric stretch of tetrazole ring together with C–H twisting out of plane. The C–H scissoring in plane is located at 967 cm⁻¹. The weak peaks less than 900 cm⁻¹ are mainly caused by the deformation of the heterocycle skeleton and the bending vibration of C–H and C–C bonds.

HOF

Heat of formation (HOF), one of the most practical parameters for experimentalists, can be used to estimate the amount of energy released or absorbed in a chemical reaction and to calculate other thermodynamic functions. DFT method has been proved to be reliable for estimating

Table 2 Thermodynamic properties of NTz-HNIW at different temperatures

T/K	$C_{p,m}^o$ /(J·mol ⁻¹ ·K ⁻¹)	S_m^o /(J·mol ⁻¹ ·K ⁻¹)	H_m^o /(kJ·mol ⁻¹)
200	331.15	668.76	39.74
298	458.44	824.75	78.50
300	460.79	827.59	79.35
400	578.26	976.72	131.48
500	673.18	1116.39	194.24
600	746.56	1245.9	265.39
700	803.00	1365.41	342.99
800	846.92	1475.62	425.57

HOF through appropriate reactions [41–43]. In this study, it has been used to compute the HOF of NTz-HNIW from the following reaction:

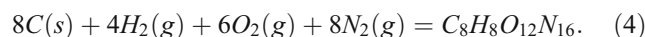


Table 1 collects the total energies (E_0) and HOFs of the species involved in the reaction. Obviously, the HOF of NTz-HNIW (755.21 kJ·mol⁻¹) is larger than those of HNIW (691.30 kJ·mol⁻¹) and TNE-HNIW (705.61 kJ·mol⁻¹), which will benefit the heat release during the detonation.

Thermodynamic properties

Thermodynamic properties, such as standard molar heat capacity $C_{p,m}^o$, entropy S_m^o , and enthalpy H_m^o , are important parameters. Based on the scaled vibrational frequencies, $C_{p,m}^o$, S_m^o , and H_m^o ranging from 200 to 800 K are obtained based on statistical thermodynamic principle and listed in Table 2. It can be seen that all thermodynamic properties increase with the temperature evidently due to the larger contributions of vibrational

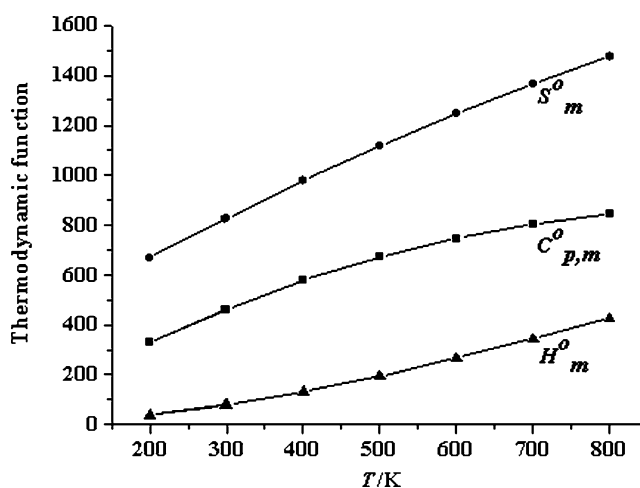
**Fig. 3** The relationships between the thermodynamic functions ($C_{p,m}^o$, S_m^o , and H_m^o) and temperature (T) for NTz-HNIW

Table 3 Unit cell parameters of the possible molecular packings of NTz-HNIW

Space groups	$P2_1/c$	$P2_12_12_1$	$P-1$	$Pbca$	$C2/c$	$Pna2_1$	$P2_1$
Z	4	4	2	8	8	4	2
$E / (\text{kJ}\cdot\text{mol}^{-1}\cdot\text{cell}^{-1})$	-358.17	-357.27	-358.05	-358.95	-358.35	-355.30	-356.33
$\rho / (\text{g}\cdot\text{cm}^{-3})$	1.998	2.125	1.994	1.990	2.062	2.040	2.043
$a / (\text{Å})$	9.87	17.66	7.72	15.07	29.59	20.20	9.37
$b / (\text{Å})$	20.59	11.75	15.64	12.56	8.06	7.83	13.47
$c / (\text{Å})$	13.03	7.84	11.16	18.34	32.91	10.71	8.92
$\alpha / (^\circ)$	90.00	90.00	123.14	90.00	90.00	90.00	90.00
$\beta / (^\circ)$	139.24	90.00	107.57	90.00	154.72	90.00	48.66
$\gamma / (^\circ)$	90.00	90.00	107.90	90.00	90.00	90.00	90.00

movement at the higher temperature. The temperature-dependent relations for $C_{p,m}^o$, S_m^o , and H_m^o are as follows and are expressed as in Fig. 3.

$$C_{p,m}^o = 10.64 + 1.78T - 9.25 \times 10^{-4}T^2 \quad (5)$$

$$S_m^o = 321.36 + 1.83T - 4.90 \times 10^{-4}T^2 \quad (6)$$

$$H_m^o = -24.55 + 0.22T + 4.27 \times 10^{-4}T^2 \quad (7)$$

The corresponding correlation coefficients are 0.9999, 0.9998, and 0.9999, respectively. These relationships and data in Table 2 will be helpful for the further studies on the other physical, chemical, and explosive properties of NTz-HNIW.

Crystal structure

Since many energetic compounds exist in crystal state, crystal structure prediction of a compound from its molecular structure is undoubtedly of great practical value. In the present work, COMPASS [46], an *ab initio* force field that can accurately predict the properties for a broad range of systems, such as HNIW and PNHAAs [47, 48], has been used to predict the possible polymorph of the title compound. Table 3 collects the unit cell parameters of the packing with the lowest energy in seven most possible space groups. One sees that the energies are in the range of $-358.95 \sim -355.30 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{cell}^{-1}$. Considering that the observed crystal structure is generally

the one with the lower energy, the title compound may crystallize in $Pbca$ symmetry. The corresponding lattice parameters are $Z=8$, $a=15.07 \text{ Å}$, $b=12.56 \text{ Å}$, $c=18.34 \text{ Å}$. The crystal density is $1.990 \text{ g}\cdot\text{cm}^{-3}$, larger than that of HNIW [45]. Figure 4 gives the $Pbca$ space group of NTz-HNIW from COPMPASS force field.

Detonation characteristics

Using the crystal density and HOF of NTz-HNIW obtained in the above sections, the detonation velocity (D), detonation pressure (P), and detonation energy (Q) were predicted using the empirical Kamlet-Jacobs equations [27]. From Table 4 we find that there are not many differences between the density, detonation velocity, detonation pressure, and detonation energy of NTz-

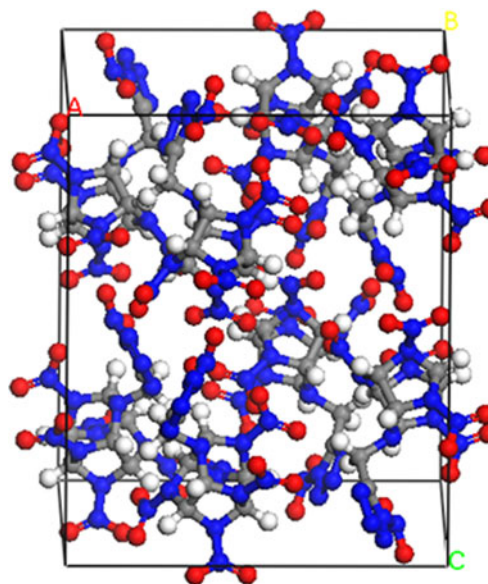
**Fig. 4** Crystal structure of NTz-HNIW in $Pbca$ space group

Table 4 Detonation performances of NTz-HNIW and HNIW

Compound	$\rho/(\text{g}\cdot\text{cm}^{-3})$	$D/(\text{km}\cdot\text{s}^{-1})$	$P/(\text{GPa})$	$Q/(\text{kJ}\cdot\text{g}^{-1})$
NTz-HNIW	1.990	9.240	40.136	-6.336
HNIW	1.970 ^a (2.040 ^b)	9.732 ^a (9.380 ^b)	44.640 ^a	-7.269 ^a (-6.234 ^c)

^a Calculated value from Ref.[45]

^b Experimental value from Ref.[1]

^c Experimental value from Ref.[50]

HNIW and HNIW. All detonation properties of NTz-HNIW satisfy the standards (i.e., $\rho=1.9 \text{ g}\cdot\text{cm}^{-3}$, $D=9.0 \text{ km}\cdot\text{s}^{-1}$, $P=40.0 \text{ GPa}$) [49] as HEDC.

Pyrolysis mechanism and thermal stability

Thermal stability is a fundamental and very important property of energetic materials. It is believed to be tightly related with the initial stages of the thermal decomposition of compounds and can be deduced on the basis of *BDEs*. At present, people have reached a consensus that nitro groups often represent the primary cause of initiation reactivity of organic polynitro compounds [51–57]. For the molecule studied here, collapse of the cage may also happen since cage compounds are always subject to a certain amount of strain energy. Therefore, two possible initial steps, i.e., the breaking of N–NO₂ bond in the side chain and C–C bond in the skeleton, are considered. NBO analysis was carried out first to find the weakest bond. The bonds with the smallest Wiberg bond index (*WBI*) among the same kind of bonds, C(1)–C(6) in the cage skeleton and N(40)–NO₂ in the side chain were found and their *BDEs* were calculated. From Table 5 we see that N(40)–NO₂ (*BDE*=156.64 kJ·mol⁻¹) is not only much weaker than other N–NO₂ bonds, but also

Table 5 Wiberg bond index, charges on nitro groups (e), and bond dissociation energy (kJ·mol⁻¹)

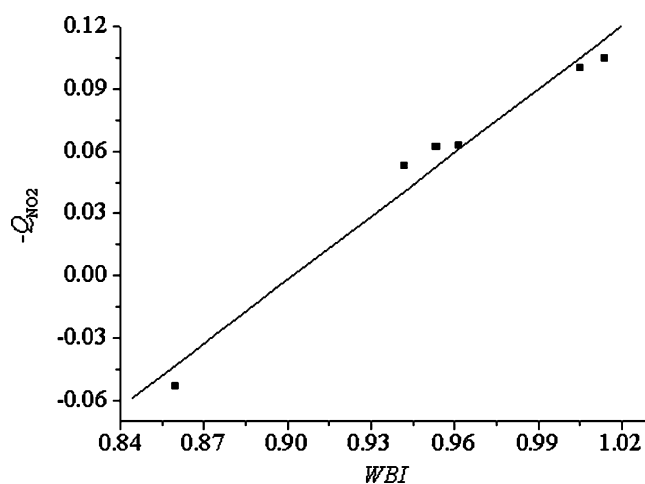
Bond	WBI	- Q_{NO_2}	BDE
N(2)–NO ₂	1.0049	0.100	/
N(5)–NO ₂	0.9535	0.062	/
N(7)–NO ₂	0.9418	0.053	/
N(10)–NO ₂	1.0137	0.105	/
N(11)–NO ₂	0.9614	0.063	/
N(40)–NO ₂	0.8598	-0.053	156.64
C(1)–C(6)	0.9205	/	238.97
C(3)–C(4)	0.9230	/	/
C(8)–C(9)	0.9220	/	/

much weaker than C(1)–C(6) (238.97 kJ·mol⁻¹). Therefore, N(40)–NO₂ breaks more easily. The *BDE* of N(40)–NO₂ is slightly smaller than that of HNIW (161.38 kJ·mol⁻¹). However, it is still large enough to suffice the stability request for HEDC (*BDE*≈80~120 kJ·mol⁻¹) suggested previously[49].

Further analysis of the NBO results, we find that there exists an essentially linear relationship between the *WBIs* of N–NO₂ bonds and the charges $-Q_{\text{NO}_2}$ on the nitro groups (Fig. 5). For example, the *WBI* of N(40)–NO₂ is the smallest, the corresponding $-Q_{\text{NO}_2}$ is also the smallest, which suggests the charges on the nitro groups can also be used to identify the strength of N–NO₂ bonds or the stability of the nitro compounds.

Summary

Theoretical investigations have been performed on a new designed polynitro cage compound 10-(1-nitro-1, 2, 3, 4-tetraazol-5-yl) methyl-2, 4, 6, 8, 12-hexanitrohexaazaisowurtzitan (NTz-HNIW) using DFT and molecular mechanics methods. IR spectrum, thermodynamic properties, and crystal structure were predicted. The most possible packing structure has the *Pbca* symmetry. HOF and detonation properties have been evaluated and results revealed that the title compound possesses very high HOF, detonation velocity, detonation pressure, and detonation energy. Stability evaluated by the *BDE* of the weakest bonds shows that the title compound also has a high thermal stability. All calculation results suggest that NTz-HNIW has comparable performances with HNIW and is a potential candidate of HEDC.

**Fig. 5** The relationship between $-Q_{\text{NO}_2}$ and *WBI* for NTz-HNIW

Acknowledgments The project is supported by the National Natural Science Foundation of China NSAF (Grant No. 11076017) and the Foundation of State Key Laboratory of Explosion Science and Technology of China (Grant No. KFJJ10-12 M)

References

- Sikder AK, Sikder N (2004) *J Hazard Mater* 112:1–15
- Millar RW, Philbin SP, Claridge RP, Hamid J (2004) *Propell Explos Pyrot* 29:81–92
- Chapman RD, Wilson WS, Fronabarger JW, Merwin LH, Ostrom GS (2002) *Thermochim Acta* 384:229–243
- Rajendra PS, Rajendar DV, Dayal TM, Jean'ne MS (2006) *Angew Chem Int Ed* 45:3584–3601
- Nielsen AT, Nissan RA, Vanderah DJ, Coon CL, Gilardi RD, George CF, Anderson JF (1990) *J Org Chem* 55:1459–1466
- Zhang MX, Eaton PE, Gilardi R, Gelber N, Iyer S, Surapaneni R (2002) *Propell Explos Pyrot* 27:1–6
- Zhang MX, Eaton PE, Gilardi R et al (2000) *Propell Explos Pyrot* 12:1143–1148
- Schulman JM, Disch RL (1984) *J Am Chem Soc* 106:1202–1204
- Xu XJ, Xiao HM, Gong XD, Ju XH, Chen ZX (2005) *J Phys Chem A* 109:11268–11274
- Stetter H, Mayer J, Schwarz M, Wolff K (1960) *Chem Ber* 93:226–230
- Sun CH, Zhao XQ, Li YC, Pang SP (2010) *Chin Chem Lett* 21:572–575
- Zhang JY, Du HC, Wang F, Gong XD, Huang YS. DFT studies on a high energy density cage compound 4-trinitroethyl-2,6,8,10,12-pentanitroheptaisowurtzitane revisions. *J Phys Chem A* (submitted 2010)
- Pagoria PF, Lee GS, Mitchell AR, Schmidt RD (2002) *Thermochim Acta* 384:187–204
- Xu XJ, Xiao HM, Ju XH, Gong XD, Zhu WH (2006) *J Phys Chem A* 110:5929–5933
- Qui L, Xiao HM, Gong XD, Ju XH, Zhu WH (2006) *J Phys Chem B* 110:3797–3807
- Xu XJ, Xiao HM, Ju XH, Gong XD, Zhao XC (2005) *J Phys Chem A* 109:11268–11274
- Qui L, Xiao HM, Zhu WH, Ju XH, Gong XD (2006) *Chin J Chem* 24:1538–1546
- Frisch MJ et al (2003) *Gaussian 03, Revision A.1*. Gaussian Inc, Pittsburgh, PA
- Materials studio 4.4. (2009) Accelrys
- Frisch MJ, Pople JA, Binkley JS (1984) Self-consistent molecular-orbital methods 25. Supplementary functions for Gaussian-basis sets. *J Chem Phys* 80:3265–3269
- Becke AD (1992) *J Chem Phys* 97:9173–9178
- Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37:785–789
- Parr RG, Yang W (1989) *Density Functional Theory of Atoms and Molecules*. Oxford University Press, Oxford
- Seminario JH (ed) (1996) *Recent developments and applications of modern density functional theory*. Elsevier, Amsterdam
- Jursic BS (1996) *THEOCHEM* 370:65–69
- Glendening ED, Reed AE, Carpenter JE, Weinhold F (1988) *NBO, Version 3.1*. Madison, WI
- Kamlet MJ, Jacobs SJ (1968) *J Chem Phys* 48:23–35
- Benson SW (1976) *Thermochemical Kinetics*. Wiley, New York
- Yao XQ, Hou XJ, Wu GS, Xu YY, Xiang HW, Jiao H, Li YW (2002) *J Phys Chem A* 106:7184–7189
- Shao J, Cheng X, Yang X (2005) *THEOCHEM* 755:127–130
- Fan XW, Ju XH, Xia QY, Xiao HM (2008) *J Hazard Mater* 151:255–260
- Chernikova NY, Belsky VK, Zorkii PM (1990) *J Struct Chem* 31:661–666
- Mighell AD, Himes VL, Rodgers JR (1983) *Acta Crystallogr A* 39:737–740
- Wilson AJC (1988) *Acta Crystallogr A* 44:715–724
- Srinivasan R (1992) *Acta Crystallogr A* 48:917–918
- Baur WH, Kassner D (1992) *Acta Crystallogr B* 48:356–369
- Wang GX, Shi CH, Gong XD, Xiao HM (2008) *THEOCHEM* 869:98–104
- Chen ZX, Xiao JM, Xiao HM, Chiu YN (1999) *J Phys Chem A* 103:8062–8066
- Zhang J, Xiao HM (2002) *J Chem Phys* 116:10674–10678
- Scott AP, Radom L (1996) *J Phys Chem* 100:16502–16513
- Jursic BS (1997) *J Chem Phys* 106:2555–2558
- Jursic BS (1997) *THEOCHEM* 391:75–81
- Jursic BS (1997) *THEOCHEM* 417:99–106
- Lide DR (ed) (2005) *CRC Handbook of Chemistry and Physics, Internet Version*
- Ghule VD, Jadhav PM, Patil RS, Radhakrishnan S, Soman T (2010) *J Phys Chem A* 114:498–503
- Sun H (1998) *J Phys Chem B* 102:7338–7364
- Xu XJ, Zhu WH, Xiao HM (2007) *J Phys Chem B* 111:2090–2097
- Zhang J, Xiao HM (2002) *Chem Phys* 116:10674–10683
- Xiao HM, Xu XJ, Qiu L (2008) *Theoretical design of high energy density materials*. Science Press, Beijing
- Simpson RL, Urtiew PA, Ornellas DL, Moody GL, Scribner KJ, Hoffman DM (1997) *Propell Explos Pyrot* 22:249–255
- Gonzalez AC, Lamon CW, McMillen DF, Golden DM (1985) *J Phys Chem* 89:4809–4814
- Owens FJ, Jayasuriya K, Abrahmsen L, Politzer P (1985) *Chem Phys Lett* 116:434–438
- Xiao HM, Fan JF, Gu ZM, Dong HS (1996) *Chem Phys* 226:15–24
- Murray JS, Politzer P (1990) In: Bulusu SN (ed) *Chemistry and Physics of Energetic Materials*. Kluwer, Dordrecht, p 175
- Michels HH, Montgomery JA (1993) *J Phys Chem* 97:6602–6606
- Pospíšil M, Vávra P, Concha MC, Murray JS, Politzer P (2011)
- Murray JS, Concha MC, Politzer P (2009) *Mol Phys* 107:89–97

A selectivity study on mTOR/PI3K α inhibitors by homology modeling and 3D-QSAR

Ting Ran · Tao Lu · Haoliang Yuan · Haichun Liu ·
Jian Wang · Weiwei Zhang · Ying Leng · Guowu Lin ·
Shulin Zhuang · Yadong Chen

Received: 23 December 2010 / Accepted: 9 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract The phosphatidylinositol-3-kinase (PI3K)/Akt/mammalian target of rapamycin (mTOR) signaling pathway plays a critical role in the regulation of cellular growth, survival and proliferation. mTOR and PI3K have attracted particular attention as cancer targets. These kinases belong to the phosphatidylinositol-3-kinase-related kinase (PIKK) family and therefore have considerable homology in their active sites. To accelerate the discovery of inhibitors with selective activity against mTOR and PI3K as cancer targets, in this work, a homology model of mTOR was developed to identify the structural divergence in the active sites between mTOR and PI3K α . Furthermore, two highly predictive comparative molecular similarity index analyses (CoMSIA) models were built based on 304 selective inhibitors docked into mTOR and PI3K α , respectively (mTOR: $q^2=0.658$, $r_{pre}^2=0.839$; PI3K α : $q^2=0.540$, $r_{pre}^2=0.719$). The results showed that steric and electrostatic fields have an important influence on selectivity towards mTOR and PI3K α —a finding consistent with the structural

divergence between the active sites. The findings may be helpful in investigating selective mTOR/PI3K α inhibitors.

Keywords mTOR · PI3K α · Selectivity · Homology modeling · 3D-QSAR

Introduction

The PI3K signaling pathway is crucial to many aspects of cell growth and survival via its regulation of widely divergent physiological processes including cell cycle progression, differentiation, transcription, translation and apoptosis [1]. Dysregulation, either through amplification of PI3K, deletion of PTEN (phosphatase and tensin homology protein) or activating mutations, has been linked closely to the development and progression of a wide range of hyper proliferative diseases and cancers [2]. This has prompted intense interest in the development of small molecule modulators of key proteins in this signaling cascade.

PI3Ks are lipid kinases that regulate cellular growth and metabolism by phosphorylation of the 3-hydroxy of phosphatidylinositol to generate phosphatidylinositol triphosphate (PIP3) [3, 4]. Among different subtypes of PI3Ks, PI3K α is the primary target for the treatment of cancers since functional loss of PTEN (the most commonly mutated tumor-suppressor gene in cancer after p53), oncogenic mutations and amplification of the PIK3CA gene (which encodes PI3K α), and over-expression of Akt have been established in many human cancers [5].

mTOR is a serine/threonine protein kinase that exists as two functional protein complexes mTORC1 and mTORC2 [6, 7]. This kinase is the main downstream component of

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1034-3) contains supplementary material, which is available to authorized users.

T. Ran · T. Lu · H. Yuan · H. Liu · J. Wang · W. Zhang · Y. Leng ·
G. Lin · Y. Chen (✉)
Laboratory of Molecular Design and Drug Discovery,
College of Basic Science, China Pharmaceutical University,
24 Tongji Xiang,
Nanjing 210009, China
e-mail: ydchen@cpu.edu.cn

S. Zhuang
Institute of Environmental Science, Zhejiang University,
388 Yuhangtang Road,
Hangzhou 310058, China

the PI3K signaling pathway and responds to a variety of extracellular signals [8]. It promotes cell growth and cell cycle progression by phosphorylating the translational regulators S6K and 4EBP [9]. An increased level of mTOR signaling has been detected in various human malignancies, thus, mTOR-specific inhibitors may demonstrate highly desirable anticancer efficacy.

The PI3K/mTOR pathway has been the subject of widespread and intense drug discovery efforts for a number of years [10, 11]. Especially mTOR and PI3K have received particular attention as cancer targets [12]. These kinases belong to the PIKK family and therefore share considerable homology in their active sites, enabling the discovery of small molecule inhibitors with varying degrees of activity against mTOR and the isoforms of PI3K [13]. Some dual-pan mTOR/PI3K inhibitors have been reported [14–16], but reports describing mTOR-selective compounds have appeared only recently [17–19]. A selectivity study on mTOR/PI3K α inhibitors would help to generate a full spectrum of compounds from the most mTOR-selective to the most PI3K-selective. Such compounds provide valuable tools with which to determine the relative roles of these kinases in cancer biology, and to facilitate the development of compounds that will have the best therapeutic potential for the treatment of cancer and other proliferative diseases [20, 21].

To accelerate the discovery of small molecule inhibitors with selective activity against mTOR and the isoforms of PI3K, a homology model of mTOR was herein developed to identify the structural divergence of the active sites compared to PI3K α [22, 23], and 3D-QSAR models of mTOR and PI3K α were developed based on a dataset of docked poses of compounds with selective activity toward mTOR and PI3K α [24]. This study thus investigated the key structural factors responsible for selectivity toward mTOR and PI3K α . The results may help in the design of selective mTOR/PI3K α inhibitors.

Materials and methods

Homology model

A homology model of mTOR was generated using the following protocol: amino acid sequences of human mTOR (accession number: P42345) and PI3K α (accession number: P42336) were retrieved from the UniProt knowledgebase (UniProtKB; <http://www.uniprot.org>). The sequence of mTOR was then taken as a query to search the Protein Data Bank (PDB; <http://www.wwpdb.org/>) using BLAST to find templates. Finally, the PI3K γ (PDB code: 3IBE) was chosen as the template for homology modeling of mTOR. The amino acid sequence of PI3K γ (accession

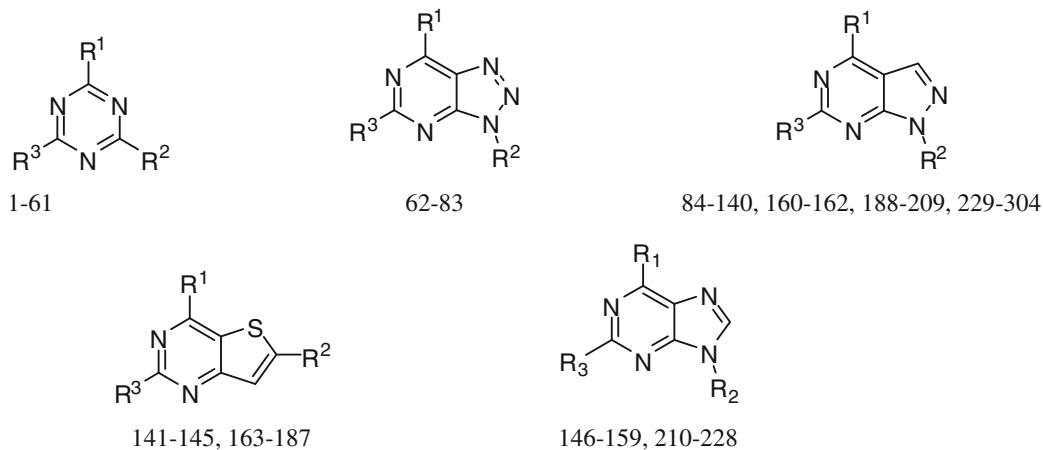
number: P48376) was also therefore retrieved from UniProtKB. mTOR, PI3K α and PI3K γ sequences were aligned using ClustalW. The resulting alignment was adjusted manually to remove gaps in uniform secondary structures. On the basis of the alignment, homology modeling was performed using the Build Homology Models module of Discovery Studio 2.5 (Accelrys; <http://accelrys.com>). In the modeling process, protein models were generated based on the spatial restraints derived from the alignment between target and template sequences, and insertions of the target sequence were dealt with as loops. Models were ranked by their probability density function (PDF) energy. The PDF scoring function mainly evaluates the imposed restraints required to force the model towards the structure of the template and to maintain good structure geometry. Models with lower PDF energy better satisfy homology restraints. If models show PDF energy values extraordinarily close to each other, the discrete optimized protein energy (DOPE) score based on statistical potentials can be used to measure the quality of the model. Finally, the best model with PDF energy (3,181.22) and DOPE score (–56,359.2) is chosen.

Energy minimization

With the objective of relieving the structural strain stemming from the replacement of non-conserved residues in the homology modeling process, energy minimization was carried out using Amber 10 (<http://ambermd.org/>). The AMBER03 force field and explicit solvent model were applied, and sodium ions were added to neutralize the charge of the system after being immersed into a rectangular truncated octahedron of TIP3P water molecules with a 10 Å box. Minimization was performed in a three-step protocol: to begin with, only solvent and ionic atoms were allowed to move in the early minimization (2,500 cycles of steepest descent and conjugate gradient minimizations); secondly, amino acid side-chains were minimized (2,500 cycles of steepest descent and conjugate gradient minimizations); eventually, the whole molecular system was optimized without any constraints (5,000 cycles of steepest descent and conjugate gradient minimizations). The quality of the refined model was evaluated with respect to the conformation of protein folds and the packing environment by Ramachandran plot and the Profile-3D program and compared to the crystal structures of PI3K α and PI3K γ .

Molecular preparation and docking

Three dimensional (3D) structures of all compounds for the 3D-QSAR study were constructed based on compound 246 in complex with PI3K γ using SYBYL-X software ([!\[\]\(ab4e2b3fc7e7887b7a72f548aa6f5e60_img.jpg\) Springer](http://</p></div><div data-bbox=)

Table 1 Structures of training set compounds

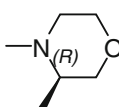
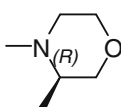
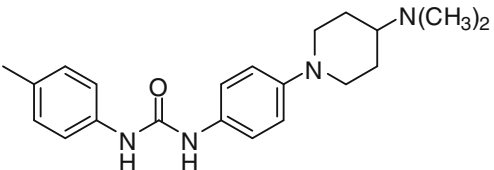
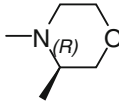
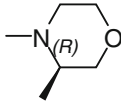
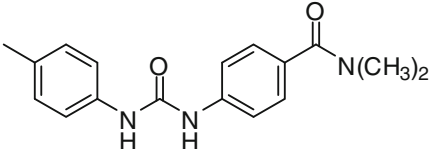
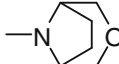
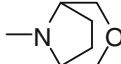
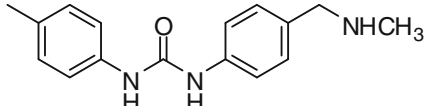
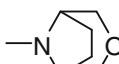
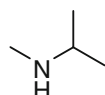
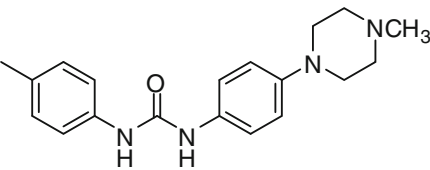
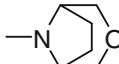
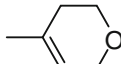
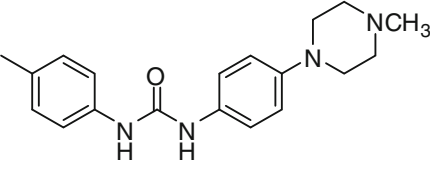
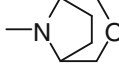
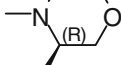
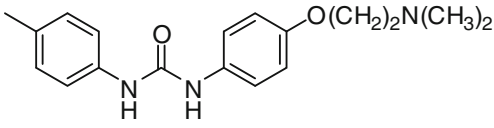
^a Comp. No.	R ¹	R ²	R ³
8*			
12			
23			
30			
34			
53			

Table 1 (continued)

67		$-\text{CH}(\text{CH}_3)_2$	
72			
73		$-\text{CH}_2\text{CH}_3$	
85			
87*			
89			
95		$-\text{CH}_2\text{CH}_2\text{OH}$	
106			
109			
112			
115			
119		$-\text{CH}_2\text{CH}_2\text{F}$	
124		$-\text{CH}_2\text{CH}_2\text{F}$	

Table 1 (continued)

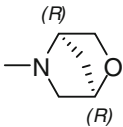
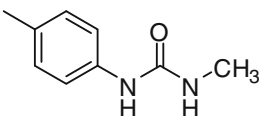
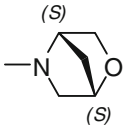
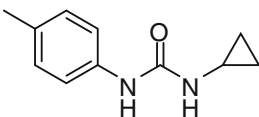
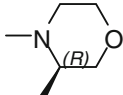
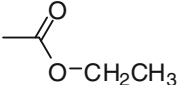
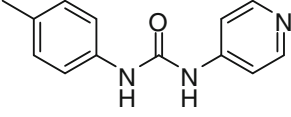
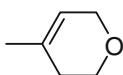
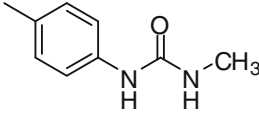
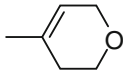
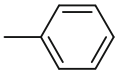
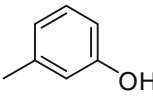
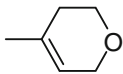
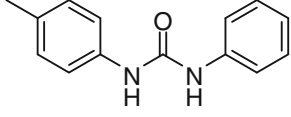
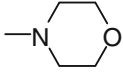
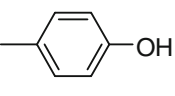
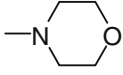
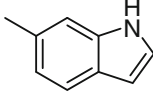
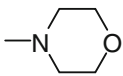
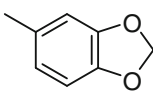
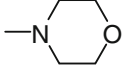
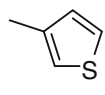
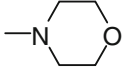
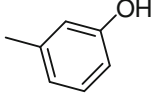
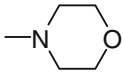
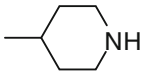
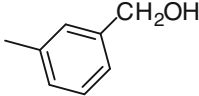
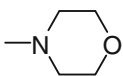
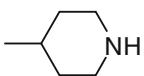
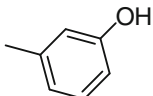
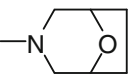
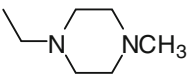
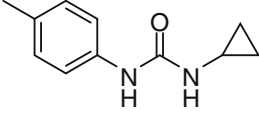
128		$-\text{CH}_2\text{CH}_2\text{F}$	
129*		$-\text{CH}_2\text{CH}_2\text{F}$	
135			
136		$-\text{CH}_2\text{CH}_2\text{F}$	
139			
143		—	
146		—	
148		—	
149		—	
150		—	
151		—	
158			
159			
164			

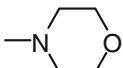
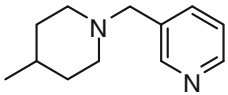
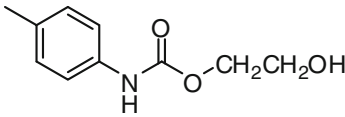
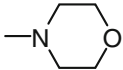
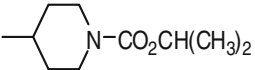
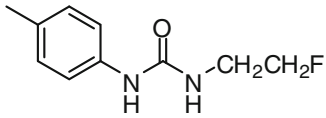
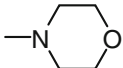
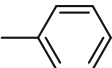
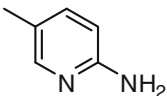
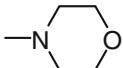
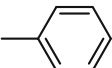
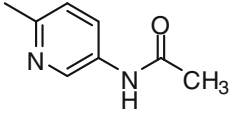
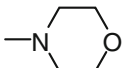
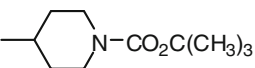
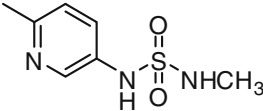
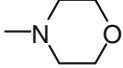
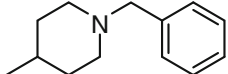
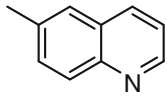
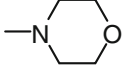
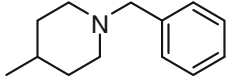
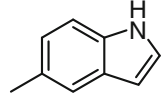
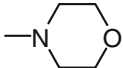
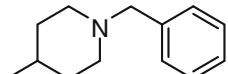
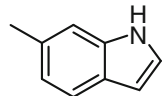
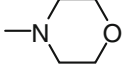
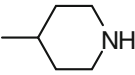
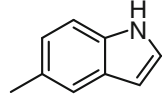
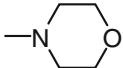
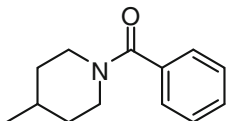
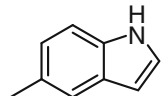
Table 1 (continued)

170			
173*		—	
176		—	
181		—	
187		—	
189		$-(\text{CH}_2)_2\text{N}(\text{CH}_3)_2$	
192		$-(\text{CH}_2)_2\text{N}(\text{CH}_3)_2$	
196		$-\text{CH}_2\text{CH}_3$	
200		$-\text{CH}_2\text{CH}_3$	
215			
223			
229*			
231			
233			

Table 1 (continued)

235			
236			
238			
240			
243			
246			
248			
253*			
254			
257			
259			
261			
262			

Table 1 (continued)

267			
270			
278			
281			
283*			
294			
298			
299			
302			
304			

tripos.com). All structures were added to the Gasteiger-Hückel charges, and energy minimization was performed using Powell gradient algorithm with the Tripos force field and a convergent threshold of a maximum deviate of $0.05 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$.

Molecular docking was performed using the Gold 3.1 docking program (<http://www.csc.fi/english/research/software/gold>). Each compound was docked into the homology model of mTOR and the crystallographic structure of PI3K α (PDB code: 2RD0), respectively. All docking calculations were undertaken in the default settings, which give best fit for flexible molecules. Early termination was unselected and no constraints were defined in the docking runs. The radius of

the active site was extended to 15 \AA while the annealing parameters were retained at 4.0 \AA for van der Waals and 2.5 \AA for hydrogen bond interactions. Default GOLDScore was employed for pose ranking.

3D-QSAR

In this study, a dataset of 304 selective ATP-competitive inhibitors was selected to construct and validate 3D-QSAR models. All compounds collected from the literature were reported by the same research group, and were divided into training (Table 1) and test (Table S1, Supplementary Material) sets according to biological activity range and

structural diversity. Since the experimental mTOR and PI3K α activities varied significantly, different training sets were built for the two cases. Inhibitory concentration values (IC₅₀) were normalized and converted to the logarithm unit of molar grade (pIC₅₀). The best docking poses of the inhibitors were aligned together into the active site as an active conformation alignment for the development of 3D-QSAR models (Fig. 1).

Comparative molecular similarity indices analysis (CoMSIA) was employed to establish 3D-QSAR models using the QSAR module in SYBYL-X software. Using a probe of a sp³ carbon atom with a charge of +1.0, five similarity indices consisting of steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor fields were calculated for each lattice with a grid of 2 Å. Partial least-squares (PLS) analysis was driven referring to the CoMSIA descriptors as explanatory variables and the calculated pIC₅₀ values as dependent variables. The mini-

imum sigma value of 2.0 kcal mol⁻¹ was set as column filter to reduce noise in the PLS analysis. The leave-one-out (LOO) algorithm was adopted in cross validation to obtain the optimal component, lowest standard error of prediction and cross-validation coefficient (q^2). The obtained optimal component was then applied in the non-cross-validation of the PLS model, resulting in a correlation coefficient (r^2). CoMSIA contour maps were displayed by interpolating pairwise products between PLS coefficients and the standard deviations of the corresponding CoMSIA descriptors.

External validation was performed by predicting the activities of compounds from the test set. The predictive R^2 (R_{pre}^2) was calculated using Eq. 1:

$$PRESS = \sum (Y_{actual} - Y_{predicted})^2$$

$$SD = \sum (Y_{actual} - Y_{mean})^2$$

$$R_{pre}^2 = 1 - PRESS/SD \quad (1)$$

Where $Y_{predicted}$, Y_{actual} , Y_{mean} are the predicted, actual and mean values of pIC₅₀. PRESS represents the predictive residual sum of squares, and SD represents the sum of the squared deviations between the observed activities of the test set and the mean activity of the training set.

Results and discussion

Homology model

In the absence of an X-ray crystallographic structure of mTOR, we developed a homology model to identify structural divergence of the active site compared to PI3K α .

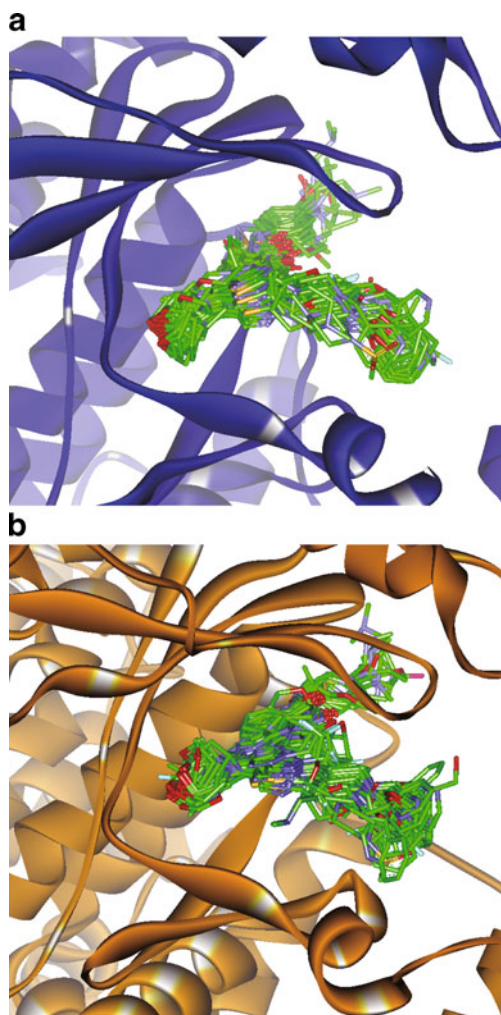


Fig. 1 Alignment of training set compounds docked within the active sites of **a** mammalian target of rapamycin (mTOR) and **b** phosphatidylinositol-3-kinase (PI3K α)

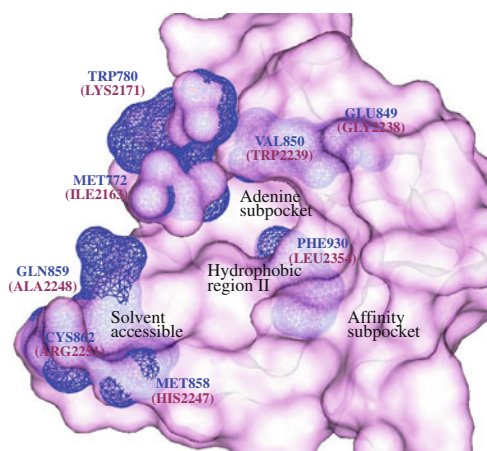


Fig. 2 Superimposition of the active sites of mTOR and PI3K α . mTOR is shown by *purple solid surface*, and the variant residues in PI3K α is shown by *blue mesh*. Residues of mTOR are labeled in *blue*, and those of PI3K α are in *purple*

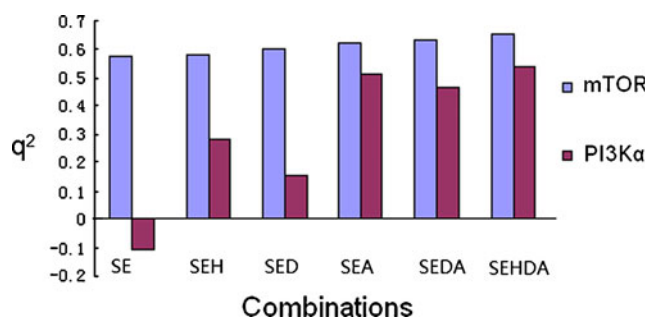


Fig. 3 Cross-validation coefficients (q^2) for combinations of different descriptors. *S* steric, *E* electrostatic, *H* hydrophobic, *D* hydrogen bond donor, *A* hydrogen bond acceptor

The multiple sequence alignment is shown in Fig. S1 (Supplementary Material). mTOR shares a conserved kinase catalytic domain with PI3K α and PI3K γ with, on average, ~17% sequence identity (~40% similarity). Generally, a sequence identity less than 30% means an unreliable alignment; however, in mTOR, the relatively low sequence identity of the kinase catalytic domain can be compensated for by a high structural similarity, particularly for the conserved residues that interact with ATP (SER2165, LYS2187, TYR2225 and VAL2240 in mTOR) and surround the ATP binding site (PRO2169, ILE2185, ILE2237 and THR2245 in mTOR) [25]. This would guarantee the production of rational protein models [26]. As expected, the constructed model exhibits reasonable conformation. An evaluation of the protein backbone shows that approximate 97% of the backbone Psi/Phi angles reside in the favorable

region according to the Ramachandran plot (Fig. S2a, Supplementary Material). Profile-3D scores calculated for the model suggest that the 3D structure has a nice compatibility with the amino acid sequence (Fig. S2b, Supplementary Material). The homology model of mTOR was overlaid onto the crystal structures of PI3K α and PI3K γ with backbone RMSDs of 1.765 Å and 0.551 Å, respectively. The overall secondary structures of the modeled mTOR and the crystal structures of the two PI3K subtypes (Fig. S3, Supplementary Material) also including those non-conserved structural domains are very similar. The results demonstrate that our mTOR homology model is well within the criteria that have been established for reliable structures and would be a reasonable starting point for further modeling and docking.

Comparison of the structures of the active sites in mTOR and PI3K α

To identify the structural divergence of the active sites between mTOR and PI3K α , the protein model of mTOR was superimposed on the crystal structure of PI3K α . Four pharmacophoric subregions within the active site were defined: (1) an adenine subpocket (hinge), (2) an “affinity” subpocket, (3) a hydrophobic region at the mouth of active site, and (4) the solvent accessible region (Fig. 2) [27]. In the adenine subpocket, although the residue that makes the key hinge hydrogen bond is conserved (VAL2240 in mTOR and VAL851 in PI3K α), some important variations were found. As shown by the protein superimposition, TRP2239

Table 2 Parameters of comparative molecular similarity indices analysis (CoMSIA) models. *mTOR* Mammalian target of rapamycin, *PI3K α* phosphatidylinositol-3-kinase. *S* steric, *E* electrostatic, *H* hydrophobic, *D* hydrogen bond donor, *A* hydrogen bond acceptor

mTOR CoMSIA							PI3K α CoMSIA					
	SE	SEH	SED	SEA	SEDA	SEHDA	SE	SEH	SED	SEA	SEDA	SEHDA
q^2 ^a	0.575	0.583	0.603	0.620	0.632	0.658	-0.107	0.279	0.156	0.517	0.468	0.540
ONC ^b	3	3	6	2	4	5	2	6	6	7	7	7
SEE ^c	0.676	0.632	0.642	0.662	0.490	0.349	0.613	0.240	0.264	0.201	0.187	0.166
r^2 ^d	0.728	0.763	0.895	0.735	0.860	0.930	0.506	0.929	0.915	0.951	0.958	0.967
F ^e	58.926	70.681	89.951	93.07	99.596	169.135	30.775	123.032	99.830	153.825	179.002	228.015
S ^F ^f	0.341	0.217	0.242	0.260	0.148	0.118	0.313	0.216	0.253	0.192	0.171	0.131
E ^F ^f	0.659	0.446	0.429	0.398	0.270	0.213	0.687	0.390	0.448	0.343	0.296	0.217
H ^F ^f		0.337				0.196		0.393				0.222
D ^F ^f			0.329		0.252	0.207			0.259		0.154	0.122
A ^F ^f				0.396	0.330	0.265				0.465	0.379	0.298

^a Cross-validated correlation coefficient after leave-one-out procedure

^b Optimal number of principal components

^c Standard error of estimate

^d Correlation coefficient

^e F-test value

^f Fraction for each field

Table 3 Actual and predicted activities for training set compounds

Compound no.	mTOR		PI3K α	
	Actual pIC ₅₀	Predicted pIC ₅₀	Actual pIC ₅₀	Predicted pIC ₅₀
8 ^a	8.699	8.693	-	-
12	9.301	9.157	5.601	5.441
23	9.051	8.834	6.646	6.706
30	8.268	7.909	6.215	6.294
34	9.155	8.757	7.387	7.451
53	8.77	8.536	6.384	6.582
67	9.244	8.84	7.979	8.004
72	8.516	8.021	8.523	8.417
73	7.678	8.05	8.097	8.117
85	9.481	9.346	6.152	6.351
87 ^a	9.62	9.257	-	-
89	9.131	9.212	6.188	6.267
95	8.545	8.661	6.426	6.274
106	7.076	7.145	5.412	5.709
109	9.319	9.204	6.169	5.687
112	9.959	10.212	6.213	6.659
115	9.208	9.245	5.255	5.138
119	9.638	9.32	8	7.888
124	9	8.722	6.712	6.724
128	7.041	7.923	5.334	5.32
129 ^a	8.721	8.777	-	-
135	9.553	9.374	6.287	6.248
136	8.699	8.299	6.699	6.552
139	7.745	7.568	7.229	7.206
143	8.143	7.964	7.553	7.704
146	6.347	6.031	5.596	5.777
148	6.824	6.421	5.633	5.695
149	5.051	5.414	5.917	5.863
150	5.409	5.575	5.623	5.767
151	5.623	5.876	6.635	6.461
158	5.143	5.078	7.237	7.131
159	6.06	6.401	7.187	7.109
164	7.796	7.824	6.058	5.854
170	9.347	9.461	8.081	8.139
173 ^a	7.495	7.226	-	-
175	7.658	7.87	5.065	5.166
181	9.469	9.243	6.489	6.407
187	7.215	7.512	8.071	8.11
189	7.886	8.369	5.505	5.461
192	8.569	9.238	6.629	6.631
196	9.523	9.519	5.667	5.574
200	8.222	7.54	5.843	5.935
215	7.292	7.027	7.26	7.411
223	5.688	6.06	7.357	7.34
229 ^a	8.018	8.691	-	-
231	8.155	8.631	6.394	6.516
233	9.42	8.599	7.638	7.094

Table 3 (continued)

Compound no.	mTOR		PI3K α	
	Actual pIC ₅₀	Predicted pIC ₅₀	Actual pIC ₅₀	Predicted pIC ₅₀
235	9.284	8.835	6.676	6.472
236	7.658	7.718	6.68	6.667
238	8.959	8.946	5.989	6.137
240	8.367	8.204	7.328	7.265
243	9.347	9.562	6.18	6.363
246	8.699	8.684	6.261	6.243
248	10.097	10.526	8.222	8.148
253 ^a	7.076	8.041	-	-
254	8.569	8.582	7.081	7.246
257	7.119	7.118	6.569	6.95
259	7.638	7.372	6.397	6.569
261	6.903	7.152	6.15	6.074
262	8.62	8.038	7.783	7.882
267	8.824	8.82	6.547	6.605
270	9.523	9.558	5.918	5.932
278	8.018	8.299	7.229	7.031
281	7.796	8.522	6.863	6.662
283 ^a	7.071	6.758	-	-
294	5.225	5.088	6.332	6.285
298	7.056	7.052	5.997	6.048
299	6.638	6.986	5.896	5.912
302	6.77	6.675	5.883	5.906
304	8.398	8.248	5.231	5.262

^a Seven compounds randomly rejected from the training set of mTOR

of mTOR is equivalent to VAL850 of PI3K α , but it occupies a larger space. This difference may lead to divergent spatial obstruction in the adenine subpockets of the two proteins. A single amino acid difference in the depth of this subpocket may also result in divergent spatial impacts (LEU2354 in mTOR vs PHE930 in PI3K α) because the LEU2354 shapes a cavity larger than that of PHE930. Additionally, the hinge GLY2238 of mTOR is equivalent to GLU849 of PI3K α . Regarding the affinity subpocket, no significant residue variation was observed because this subpocket is comprised of conserved residues (LEU2192, ASP2195, VAL2227, and ILE2237 in mTOR, not shown in Fig. 2). However, in the hydrophobic region, ILE2163 of mTOR is equivalent to MET772 of PI3K α , which possesses a bulkier side-chain and leads to larger spatial obstruction. The case is the same for LYS2171 of mTOR, which is equivalent to TRP780 of PI3K α . In addition, in the solvent accessible region, there are several non-conserved residues (such as HIS2247, ALA2248 and ARG2251 in mTOR corresponding to MET858, GLN859 and CYS862 in PI3K α , respectively) that may interact with inhibitors. For further discussion, the region in which these

residues are localized was termed the “outside subsite”. Altogether, comparative analyses of the active site revealed some important structural differences that are potentially helpful in explaining the selectivity of inhibitors against mTOR and PI3K α .

3D-QSAR

CoMSIA model

In order to further investigate the structural divergence responsible for selectivity, we performed a 3D-QSAR study using docking-based alignment. A dataset of 304 inhibitors with known inhibitory activities toward mTOR and PI3K α was used for the 3D-QSAR analyses. As shown in Table 1, the compounds of the dataset fall into five categories with different scaffolds: triazine [28–30], triazolopyrimidine [31], purine [32, 33], thienopyrimidine [34, 35] and pyrazolopyrimidine [17, 18, 36–39]. The biological activities of the inhibitors are given as pIC₅₀. The pIC₅₀ values of the dataset span 5.118 log units for mTOR and 4.079 log units for PI3K α . This fulfills the prerequisite of a spread of at least

three orders of magnitude to ensure the development of statistically significant 3D-QSAR models. In this work, 70 compounds were selected as the training set for mTOR on account of the fact that the test set compounds represent structural diversity and a range of biological activity similar to those of the training set. However, since the activities against PI3K α cover a narrower range, a smaller training set of PI3K α was generated through randomly rejecting 7 compounds from the training set of mTOR; 234 compounds were selected as the test set to validate the QSAR model.

In the CoMSIA study, computation of six combinations of different fields was performed to determine the best predictive CoMSIA model (Fig. 3). As a result, the combination of all five fields gave the highest statistics (mTOR: $q^2=0.658$, $r^2=0.930$; PI3K α : $q^2=0.540$, $r^2=0.967$) (Table 2), which indicates that all five features are important for activity. Subsequently, the predictive power of the best CoMSIA model was evaluated by estimating the binding affinities of the training set and test set compounds, respectively. As a result, for both proteins, statistically significant models were obtained. The predicted vs actual activity values for the training set compounds are shown in Table 3. The predicted values do not deviate significantly from the actual values (generally by less than one logarithmic unit). The predicted vs actual activity values for the training set compounds are depicted graphically in Fig. 4. In the activity prediction of the test set compounds, the calculation of R_{pre}^2 shows values of 0.839 and 0.719 for mTOR and PI3K α models, respectively, indicating the very high predictability of the models (Fig. S4, Table S2, Supplementary Material).

CoMSIA contour maps

The results of CoMSIA analyses were visualized as 3D contour maps aligned into the active site. Compound 111—one of the most selective mTOR/PI3K α inhibitors—was used as a reference in the contour maps, and maps of mTOR and PI3K α were compared to identify the key field responsible for the selectivity of mTOR/PI3K α targeting. It is worth noting that the CoMSIA field derived from docking-based active conformation alignment is a joint product considering both ligands and the active site. Thus, the maps will be discussed in terms of the pharmacophoric regions of the active site.

Regarding the steric contour maps in the adenine subpocket (Fig. 5a1/2), a green contour representing a sterically favored field is localized around the bridged morpholine in the mTOR plot, whereas a yellow contour representing unfavorable steric field occurs in the corresponding area of the PI3K α plot. This difference illustrates that bulky groups would be beneficial for mTOR but not for PI3K α targeting. For example, compounds 118, 120, 121, 125 and 126, with bridged morpholinyl groups, show better selectivity against mTOR than compounds 117, 119, and 124 with no bridged

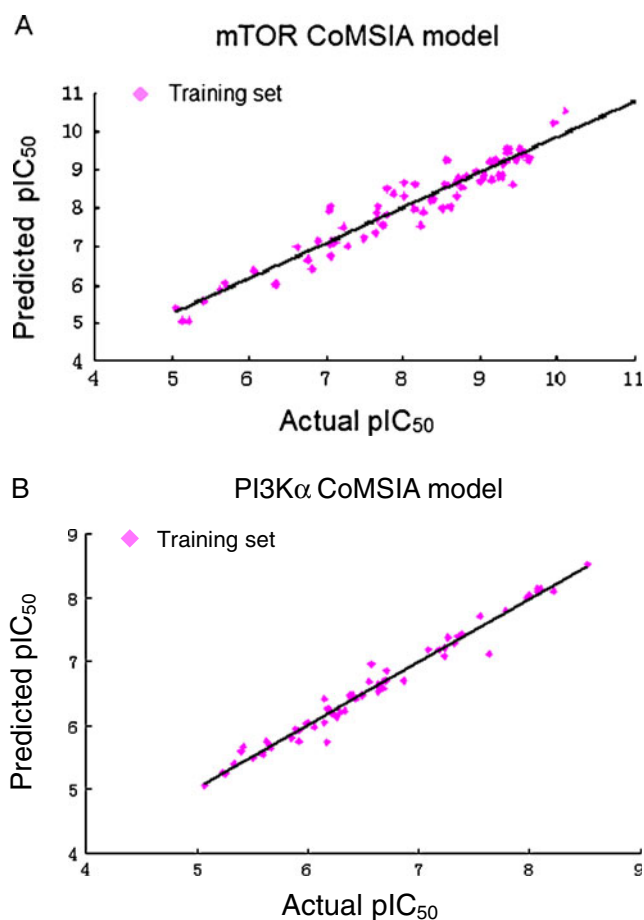
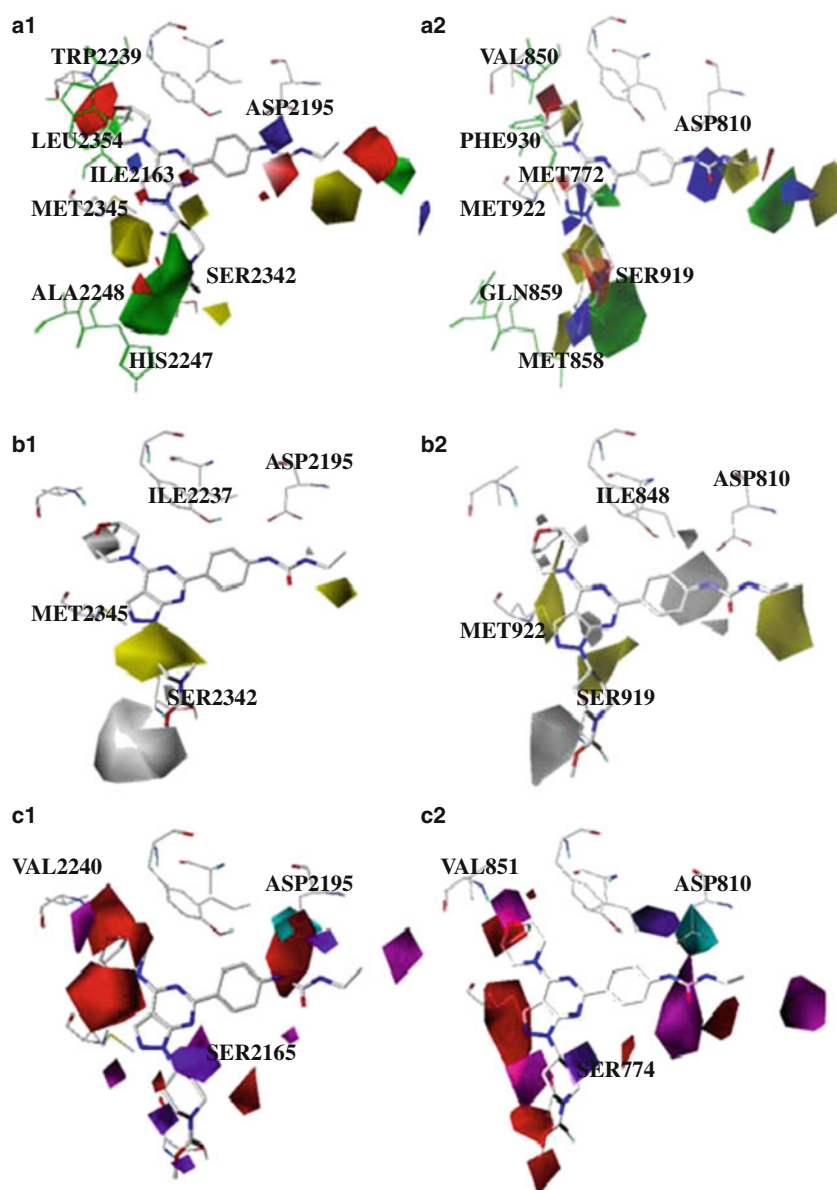


Fig. 4 **a** Plot of predicted vs actual pIC_{50} for training set compounds according to the mTOR comparative molecular similarity indices analysis (CoMSIA) model. **b** Plot of predicted vs actual pIC_{50} for training set compounds according to the PI3K α CoMSIA model

ring on the morpholinyl group. The bulky bridged group may make a sterically unfavorable interaction, which corresponds to the spatial obstructions caused by residue variations in the adenine subpocket (TRP2239 and LEU2354 in mTOR vs VAL850 and PHE930 in PI3K α). However, in the affinity subpocket, the yellow contour in the mTOR plot is also present in the PI3K α plot. This indicates that conformational restraints exist in the subpocket of both proteins (cf. activities of compound 261 and those of compounds 235, 255–260, and 262–265). The case is similar for the green contour representing the sterically favored field near the affinity subpocket, which explains the different biological activities among compounds 33–42. In the hydrophobic region, the common yellow contour, which surrounds the piperidinyl group of the reference compound, explains the relatively low activity of compound 275–282 with its bulky phenyl group against both proteins. These compounds may make sterically disfavored interactions with conserved residues such as SER2165, THR2245, SER2342 and MET2345 in mTOR (corresponding to SER774, THR856, SER919 and MET922 in PI3K α) at the mouth of

Fig. 5 CoMSIA STDEV*-COEFF contour maps based on compounds 111 as a reference. Compound 111 is shown in *stick* representation. Residues in the active site are shown as *lines*. Contours in mTOR are shown in **a1**, **b1** and **c1**, and contours in PI3K α are shown in **a2**, **b2** and **c2**. **a** Steric field: *green contour* sterically favored field, *yellow contour* sterically disfavored field. Electrostatic field: *red contour* electronegative field, *blue contour* electropositive field. **b** Hydrophobic field: *yellow contour* hydrophobic field, *white contour* hydrophilic contour. **c** Hydrogen bond field: *Cyan contour* hydrogen bond donor favored field, *blue contour* hydrogen bond donor disfavored field. *Purple contour* Hydrogen bond acceptor favored field, *brown contour* hydrogen bond acceptor disfavored field



the active site. This suggests that the residue differences (ILE2163 in mTOR vs MET772 in PI3K α) at the mouth may not have much influence on the selectivity of inhibitors. In the solvent accessible region, the green contour around the terminal methyl, indicating that the bulky substituent is favorable (cf. activities between compound 235 and compound 230, 236), is the same in both plots.

With respect to the electrostatic contour maps in Fig. 5a1/2, a red contour representing electronegative field appears in the adenine subpocket in both the mTOR and PI3K α plots. It is associated with the hinge hydrogen bond interaction engaged by VAL2240 in mTOR and VAL851 in PI3K α . Around the affinity subpocket, contour maps in mTOR and PI3K α plots are also similar. In the case of mTOR, the electropositive contour presented in blue and the electronegative contour in red correspond to the positive

LYS2187 and the negative ASP2195 and GLN2194, respectively (see the activity trend of compound 232–233, 267–268, and 251–253). As far as the electrostatic field in the hydrophobic region is concerned, fragmented contours around the scaffold ring of the reference compound are localized in a conserved region formed by LEU2185, PRO2169, MET2345 and ILE2356 of mTOR (corresponding to ILE800, PRO778, MET922, and ILE932 in PI3K α). However, electrostatic maps in the solvent accessible region are quite different. The most important difference is the map at the outside subsite. There is a red contour in the mTOR plot, but a blue one in the PI3K α plot. This difference could be responsible for mTOR/PI3K α targeting selectivity.

The hydrophobic maps in Fig. 5b1/2 are similar. For instance, the white contour representing hydrophobic disfavored field corresponds to the common hinge hydrogen bond

region in the adenine subpocket. In the affinity subpocket, the white contour corresponds to the region of the ureal moiety of the reference compound, which interacts with LYS2187, ASP2195 of mTOR and LYS802, ASP810 of PI3K α . The neighboring yellow contour may correspond to the hydrophobic interactions with LEU2192, VAL2227, and ILE2237 around the affinity subpocket in the case of mTOR. In the hydrophobic region, the hydrophobic favored contours presented in yellow are flanked by the conserved SER2165, THR2245, SER2342 and MET2345 in mTOR and SER774, THR856, SER919 and MET922 in PI3K α at the mouth of the active site.

As shown in Fig. 5c1/2, the patterns of hydrogen bond contour maps are also similar. The purple contour representing hydrogen bond acceptor favored field in the adenine subpocket suggests a conserved hydrogen bond interaction with VAL2240 of mTOR (corresponding to VAL851 in PI3K α). The purple and cyan contours in the affinity subpocket suggest hydrogen bond interactions of the ureal group of the reference compound with ASP2195, LYS2187 and GLN2345 in mTOR (comparing activities among compounds 232, 234, 267 and 233). In addition, hydrogen bond acceptor favored purple contours around the piperidinyl of the reference compound may correspond to the conserved SER774 of PI3K α (corresponding to SER2165 in mTOR) reported to be involved only in hydrogen bond interaction in the hydrophobic region [40].

Conclusions

In this work, molecular modeling methods were employed to guide study on the selectivity of mTOR/PI3K α targeting. In the absence of the crystal structures of mTOR, the protein structure of mTOR kinase was constructed by homology modeling. The structural divergence of the active site between the homology model of mTOR and the crystal structure of PI3K α were analyzed and compared; the results revealed significant differences in the residues in the adenine subpocket and the solvent accessible region. Furthermore, a 3D-QSAR study based on 304 selective inhibitors was carried out using docking-based active conformation alignment for both mTOR and PI3K α . Two highly predictive CoMSIA models were constructed (mTOR: $q^2=0.658$, $r_{pre}^2=0.839$; PI3K α : $q^2=0.540$, $r_{pre}^2=0.719$). Comparative analyses of the two CoMSIA models reveal that the steric field in the adenine subpocket and the electrostatic field in the solvent accessible region may have a critical impact on selectivity. The results agree well with the structural divergence in the two regions of the active sites of the two target proteins. These findings may provide insights that will aid discovery of novel highly selective mTOR/PI3K α inhibitors.

Acknowledgments The work was supported financially by the Natural Science Foundation of Jiangsu Province (BK2009303), the National Natural Science Foundation of China (30973609), the National Major Science and Technology Project of China (Innovation and Development of New Drugs) (No. 2009ZX09501-003), and Qing Lan Project of Jiangsu Province, China.

References

- Marone R, Cmiljanovic V, Giese B, Wymann MP (2008) Targeting phosphoinositide 3-kinase: moving towards therapy. *Biochim Biophys Acta* 1784:159–185
- Liu P, Cheng H, Roberts TM, Zhao JJ (2009) Targeting the phosphoinositide 3-kinase pathway in cancer. *Nat Rev Drug Discov* 8:627–644
- Katso R, Okkenhaug K, Ahmadi K, White S, Timms J, Waterfield MD (2001) Cellular function of phosphoinositide 3-kinases: implications for development, homeostasis, and cancer. *Annu Rev Cell Dev Biol* 17:615–675
- Anderson KE, Jackson SP (2003) Class I phosphoinositide 3-kinases. *Int J Biochem Cell Biol* 35:1028–1033
- Oda K, Okada J, Timmerman L, Rodriguez-Viciana P, Stokoe D, Shoji K, Taketani Y, Kuramoto H, Knight ZA, Shokat KM, McCormick F (2008) PIK3CA cooperates with other phosphatidylinositol 3'-kinase pathway mutations to effect oncogenic transformation. *Cancer Res* 68:8127–8136
- Dunlop EA, Tee AR (2009) Mammalian target of rapamycin complex 1: signalling inputs, substrates and feedback mechanisms. *Cell Signal* 21:827–835
- Sparks CA, Guertin DA (2010) Targeting mTOR: prospects for mTOR complex 2 inhibitors in cancer therapy. *Oncogene* 29:3733–3744
- O'Reilly KE, Rojo F, She QB, Solit D, Mills GB, Smith D, Lane H, Hofmann F, Hicklin DJ, Ludwig DL, Baselga J, Rosen N (2006) mTOR inhibition induces upstream receptor tyrosine kinase signaling and activates Akt. *Cancer Res* 66:1500–1508
- Acosta-Jaquez HA, Keller JA, Foster KG, Ekim B, Soliman GA, Feener EP, Ballif BA, Fingar DC (2009) Site-specific mTOR phosphorylation promotes mTORC1-mediated signaling and cell growth. *Mol Cell Biol* 29:4308–4324
- Wu P, Liu T, Hu Y (2009) PI3K inhibitors for cancer therapy: what has been achieved so far? *Curr Med Chem* 16:916–930
- LoPiccolo J, Blumenthal GM, Bernstein WB, Dennis PA (2008) Targeting the PI3K/Akt/mTOR pathway: effective combinations and clinical considerations. *Drug Resist Updat* 11:32–50
- Brachmann S, Fritsch C, Maira SM, Garcia-Echeverria C (2009) PI3K and mTOR inhibitors: a new generation of targeted anticancer agents. *Curr Opin Cell Biol* 21:194–198
- Lempiainen H, Halazonetis TD (2009) Emerging common themes in regulation of PIKKs and PI3Ks. *EMBO J* 28:3067–3073
- Heffron TP, Berry M, Castaneda G, Chang C, Chuckowree I, Dotson J, Folkes A, Gunzner J, Lesnick JD, Lewis C, Mathieu S, Nonomiya J, Olivero A, Pang J, Peterson D, Salphati L, Sampath D, Sideris S, Sutherland DP, Tsui V, Wan NC, Wang S, Wong S, Zhu BY (2010) Identification of GNE-477, a potent and efficacious dual PI3K/mTOR inhibitor. *Bioorg Med Chem Lett* 20:2408–2411
- Maira SM, Stauffer F, Brueggen J, Furet P, Schnell C, Fritsch C, Brachmann S, Chene P, De Pover A, Schoemaker K, Fabbro D, Gabriel D, Simonen M, Murphy L, Finan P, Sellers W, Garcia-Echeverria C (2008) Identification and characterization of NVP-BE2255, a new orally available dual phosphatidylinositol 3-kinase/mammalian target of rapamycin inhibitor with potent in vivo antitumor activity. *Mol Cancer Ther* 7:1851–1863

16. Schnell CR, Stauffer F, Allegrini PR, O'Reilly T, McSheehy PM, Dartois C, Stumm M, Cozens R, Littlewood-Evans A, Garcia-Echeverria C, Maira SM (2008) Effects of the dual phosphatidylinositol 3-kinase/mammalian target of rapamycin inhibitor NVP-BE235 on the tumor vasculature: implications for clinical imaging. *Cancer Res* 68:6598–6607
17. Curran KJ, Verheijen JC, Kaplan J, Richard DJ, Toral-Barza L, Hollander I, Lucas J, Ayrál-Kaloustian S, Yu K, Zask A (2010) Pyrazolopyrimidines as highly potent and selective, ATP-competitive inhibitors of the mammalian target of rapamycin (mTOR): optimization of the 1-substituent. *Bioorg Med Chem Lett* 20:1440–1444
18. Nowak P, Cole DC, Brooijmans N, Bursavich MG, Curran KJ, Ellingboe JW, Gibbons JJ, Hollander I, Hu Y, Kaplan J, Malwitz DJ, Toral-Barza L, Verheijen JC, Zask A, Zhang WG, Yu K (2009) Discovery of potent and selective inhibitors of the mammalian target of rapamycin (mTOR) kinase. *J Med Chem* 52:7081–7089
19. Tsou HR, MacEwan G, Birnberg G, Grosu G, Bursavich MG, Bard J, Brooijmans N, Toral-Barza L, Hollander I, Mansour TS, Ayrál-Kaloustian S, Yu K (2010) Discovery and optimization of 2-(4-substituted-pyrrolo[2,3-b]pyridin-3-yl)methylene-4-hydroxybenzofuran-3(2H)-ones as potent and selective ATP-competitive inhibitors of the mammalian target of rapamycin (mTOR). *Bioorg Med Chem Lett* 20:2321–2325
20. Gibbons JJ, Abraham RT, Yu K (2009) Mammalian target of rapamycin: discovery of rapamycin reveals a signaling pathway important for normal and cancer cell growth. *Semin Oncol* 36 (Suppl 3):S3–S17
21. Yu K, Toral-Barza L, Shi C, Zhang WG, Lucas J, Shor B, Kim J, Verheijen J, Curran K, Malwitz DJ, Cole DC, Ellingboe J, Ayrál-Kaloustian S, Mansour TS, Gibbons JJ, Abraham RT, Nowak P, Zask A (2009) Biochemical, cellular, and in vivo activity of novel ATP-competitive and selective inhibitors of the mammalian target of rapamycin. *Cancer Res* 69:6232–6240
22. Malagu K, Duggan H, Menear K, Hummersone M, Gomez S, Bailey C, Edwards P, Drzewiecki J, Leroux F, Quesada MJ, Hermann G, Maine S, Molyneux CA, Le Gall A, Pullen J, Hickson I, Smith L, Maguire S, Martin N, Smith G, Pass M (2009) The discovery and optimisation of pyrido[2,3-d]pyrimidine-2,4-diamines as potent and selective inhibitors of mTOR kinase. *Bioorg Med Chem Lett* 19(20):5950–5953
23. Cheng H, Bagrodia S, Bailey S, Edwards M, Hoffman J, Hu Q, Kania R, Knighton DR, Marx MA, Ninkovic S, Sun S, Zhang E (2010) Discovery of the highly potent PI3K/mTOR dual inhibitor PF-04691502 through structure based drug design. *Med Chem Commun* 1:139–144
24. Ortore G, Di Colo F, Martinelli A (2009) Docking of hydroxamic acids into HDAC1 and HDAC8: a rationalization of activity trends and selectivities. *J Chem Inf Model* 49:2774–2785
25. Sturgill TW, Hall MN (2009) Activating mutations in TOR are in similar structures as oncogenic mutations in PI3K α . *ACS Chem Biol* 4:999–1015
26. Blair LP, Tackett AJ, Raney KD (2009) Development and evaluation of a structural model for SF1B helicase Dda. *Biochemistry* 48:2321–2329
27. Berndt A, Miller S, Williams O, Le DD, Houseman BT, Pacold JJ, Gorrec F, Hon W-C, Ren P, Liu Y, Rommel C, Gaillard P, Ruckle T, Schwarz MK, Shokat KM, Shaw JP, Williams RL (2010) The p110[δ] structure: mechanisms for selectivity and potency of new PI(3)K inhibitors. *Nat Chem Biol* 6:117–124
28. Verheijen JC, Richard DJ, Curran K, Kaplan J, Yu K, Zask A (2010) 2-Arylureidophenyl-4-(3-oxa-8-azabicyclo[3.2.1]octan-8-yl)triazines as highly potent and selective ATP competitive mTOR inhibitors: optimization of human microsomal stability. *Bioorg Med Chem Lett* 20:2648–2653
29. Richard DJ, Verheijen JC, Yu K, Zask A (2010) Triazines incorporating (R)-3-methylmorpholine are potent inhibitors of the mammalian target of rapamycin (mTOR) with selectivity over PI3K α . *Bioorg Med Chem Lett* 20:2654–2657
30. Zask A, Verheijen JC, Richard DJ, Kaplan J, Curran K, Toral-Barza L, Lucas J, Hollander I, Yu K (2010) Discovery of 2-ureidophenyltriazines bearing bridged morpholines as potent and selective ATP-competitive mTOR inhibitors. *Bioorg Med Chem Lett* 20:2644–2647
31. Dehnhardt CM, Venkatesan AM, Delos Santos E, Chen Z, Santos O, Ayrál-Kaloustian S, Brooijmans N, Mallon R, Hollander I, Feldberg L, Lucas J, Chaudhary I, Yu K, Gibbons J, Abraham R, Mansour TS (2010) Lead optimization of N-3-substituted 7-morpholinotriazolopyrimidines as dual phosphoinositide 3-kinase/mammalian target of rapamycin inhibitors: discovery of PKI-402. *J Med Chem* 53:798–810
32. Gilbert AM, Nowak P, Brooijmans N, Bursavich MG, Dehnhardt C, Santos ED, Feldberg LR, Hollander I, Kim S, Lombardi S, Park K, Venkatesan AM, Mallon R (2010) Novel purine and pyrazolo[3,4-d]pyrimidine inhibitors of PI3 kinase- α : hit to lead studies. *Bioorg Med Chem Lett* 20:636–639
33. Venkatesan AM, Dehnhardt CM, Chen Z, Santos ED, Dos Santos O, Bursavich M, Gilbert AM, Ellingboe JW, Ayrál-Kaloustian S, Khafizova G, Brooijmans N, Mallon R, Hollander I, Feldberg L, Lucas J, Yu K, Gibbons J, Abraham R, Mansour TS (2010) Novel imidazolopyrimidines as dual PI3-Kinase/mTOR inhibitors. *Bioorg Med Chem Lett* 20:653–656
34. Kaplan J, Verheijen JC, Brooijmans N, Toral-Barza L, Hollander I, Yu K, Zask A (2010) Discovery of 3,6-dihydro-2 H-pyran as a morpholine replacement in 6-aryl-1 H-pyrazolo[3,4-d]pyrimidines and 2-arylthieno[3,2-d]pyrimidines: ATP-competitive inhibitors of the mammalian target of rapamycin (mTOR). *Bioorg Med Chem Lett* 20:640–643
35. Verheijen JC, Yu K, Toral-Barza L, Hollander I, Zask A (2010) Discovery of 2-arylthieno[3,2-d]pyrimidines containing 8-oxa-3-azabi-cyclo[3.2.1]octane in the 4-position as potent inhibitors of mTOR with selectivity over PI3K. *Bioorg Med Chem Lett* 20:375–379
36. Zask A, Verheijen JC, Curran K, Kaplan J, Richard DJ, Nowak P, Malwitz DJ, Brooijmans N, Bard J, Svenson K, Lucas J, Toral-Barza L, Zhang WG, Hollander I, Gibbons JJ, Abraham RT, Ayrál-Kaloustian S, Mansour TS, Yu K (2009) ATP-competitive inhibitors of the mammalian target of rapamycin: design and synthesis of highly potent and selective pyrazolopyrimidines. *J Med Chem* 52:5013–5016
37. Verheijen JC, Richard DJ, Curran K, Kaplan J, Lefever M, Nowak P, Malwitz DJ, Brooijmans N, Toral-Barza L, Zhang WG, Lucas J, Hollander I, Ayrál-Kaloustian S, Mansour TS, Yu K, Zask A (2009) Discovery of 4-morpholino-6-aryl-1 H-pyrazolo[3,4-d]pyrimidines as highly potent and selective ATP-competitive inhibitors of the mammalian target of rapamycin (mTOR): optimization of the 6-aryl substituent. *J Med Chem* 52:8010–8024
38. Zask A, Kaplan J, Verheijen JC, Richard DJ, Curran K, Brooijmans N, Bennett EM, Toral-Barza L, Hollander I, Ayrál-Kaloustian S, Yu K (2009) Morpholine derivatives greatly enhance the selectivity of mammalian target of rapamycin (mTOR) inhibitors. *J Med Chem* 52:7942–7945
39. Richard DJ, Verheijen JC, Curran K, Kaplan J, Toral-Barza L, Hollander I, Lucas J, Yu K, Zask A (2009) Incorporation of water-solubilizing groups in pyrazolopyrimidine mTOR inhibitors: discovery of highly potent and selective analogs with improved human microsomal stability. *Bioorg Med Chem Lett* 19:6830–6835
40. Zvelebil MJ, Waterfield MD, Shuttleworth SJ (2008) Structural analysis of PI3-kinase isoforms: identification of residues enabling selective inhibition by small molecule ATP-competitive inhibitors. *Arch Biochem Biophys* 477:404–410

Identification of novel, less toxic PTP-LAR inhibitors using in silico strategies: pharmacophore modeling, SADMET-based virtual screening and docking

Dara Ajay · M. Elizabeth Sobhia

Received: 1 December 2010 / Accepted: 10 March 2011 / Published online: 27 April 2011

© Springer-Verlag 2011

Abstract Human leukocyte antigen-related (PTP-LAR) is a receptor-like transmembrane phosphatase and a potential target for diabetes, obesity and cancer. In the present study, a sequence of in silico strategies (pharmacophore mapping, a 3D database searching, SADMET screening, and docking and toxicity studies) was performed to identify eight novel nontoxic PTP-LAR inhibitors. Twenty different pharmacophore hypotheses were generated using two methods; the best (hypothesis 2) consisted of three hydrogen-bond acceptor (A), one ring aromatic (R), and one hydrophobic aliphatic (Z) features. This hypothesis was used to screen molecules from several databases, such as Specs, IBS, MiniMaybridge, NCI, and an in-house PTP inhibitor database. In order to overcome the general bioavailability problem associated with phosphatases, the hits obtained were filtered by Lipinski's rule of five and SADMET properties and validated by molecular docking studies using the available crystal structure 1LAR. These docking studies suggested the ligand binding pattern and interactions required for LAR inhibition. The docking analysis also revealed that sulfonylurea derivatives with an isoquinoline or naphthalene scaffold represent potential LAR drugs. The screening protocol was further validated using ligand pharmacophore mapping studies, which showed that the abovementioned interactions are indeed crucial and that the

screened molecules can be presumed to possess potent inhibitory activities.

Keywords Human leukocyte antigen-related phosphatase (PTP-LAR) · Pharmacophore model · SADMET-based virtual screening · Docking · Derek toxicity studies

Abbreviations

PTP-LAR	Human leukocyte antigen-related phosphatase
PTPs	Protein tyrosine phosphatases
PTKs	Protein tyrosine kinases
IR	Insulin receptor

Introduction

Protein tyrosine phosphatases (PTPs) are a group of signal transduction enzymes which are responsible for the dephosphorylation of tyrosine-phosphorylated proteins. PTPs, together with protein tyrosine kinases (PTKs), control the level of cellular protein tyrosine phosphorylation [1–3]. Protein tyrosine phosphorylation is a major cellular post-translational modification which regulates signal transduction in cellular growth, the cell cycle, cell–cell communication, cell migration, gene transcription, ion-channel activity, differentiation, metabolism and immune response [4–7]. The protein kinases and phosphatases work in dynamic equilibrium, and exert both positive and negative regulatory effects on various signaling pathways. PTKs are responsible for protein tyrosine phosphorylation, whereas PTPs dephosphorylate cellular protein. Based on the conserved fingerprint sequence ([I/V] HCXAGXXR [S/T] G), PTPs are classified into different families [8–10]. This sequence corresponds to the

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1037-0) contains supplementary material, which is available to authorized users.

D. Ajay · M. E. Sobhia (✉)
Department of Pharmacoinformatics, National Institute
of Pharmaceutical Education and Research (NIPER),
S.A.S. Nagar,
Punjab 160062, India
e-mail: mesophia@nipер.ac.in

catalytic site containing residues such as cysteine, glutamine, aspartic acid and serine, which are essential for the dephosphorylation of the phosphotyrosine residue [11, 12]. The three major subfamilies of PTPs are classical, dual-specific and low molecular weight phosphatases [13–15]. The classical and low molecular weight phosphatases strictly target the proteins containing phosphorylated tyrosine (pTyr), whereas the dual-specific phosphatases target pTyr as well as pSer and pThr [16, 17].

Among PTPs, human leukocyte antigen-related phosphatase (LAR) is a receptor-like classical transmembrane phosphatase which acts as a negative regulator of multiple receptor tyrosine kinases [18–22]. PTP-LAR is embedded in the phospholipid membrane of the cell, and it is divided into an extracellular structure and an intracellular structure. Its extracellular structure consists of three immunoglobulin-like domains and eight fibronectin type III-like domains, while its intracellular structure consists of two tandem phosphatase catalytic domains, a membrane proximal domain (D1) and a membrane distal domain (D2) [8, 23]. Several *in vitro* studies of LAR suggest that both the D1 and the D2 domains have very similar structural characteristics. These studies also revealed that the conserved cysteine residue present in the signature motif of the D1 domain is required for catalytic activity, whereas the D2 domain has a regulatory function (Fig. 1) [24, 25].

Several biochemical and pharmacological studies indicate that LAR is a potential target for the treatment of type 2 diabetes and obesity complications [26–29]. LAR is widely detected in various insulin-sensitive tissues such as muscle tissues, the liver and adipocytes. This emphasizes the importance of LAR in insulin receptor signal transduction [30–32]. *In vitro* studies reveal that LAR not only exhibits physical association with the insulin receptor (IR) but also decreases IR autophosphorylation and kinase activity by 47% [33, 34]. When compared with PTP1B, LAR has been shown to deactivate IR kinase 3.1 times more rapidly, by preferentially dephosphorylating the IR Tyr-1150 residue that is crucial to receptor activity [35]. The overexpression of human LAR in muscle tissues of transgenic mice caused insulin resistance due to the dephosphorylation of specific regulatory phosphotyrosines on IR substrate proteins [36, 37]. Genetic studies of LAR single-strand conformation polymorphism (SSCP) using automated DNA sequencing analysis show that LAR reduces the risk of obesity and insulin resistance [38]. Many biological studies highlight the importance of LAR as a negative regulator in insulin signaling [39–41].

However, the availability of potent LAR inhibitors that can act as drug molecules in the treatment of insulin-related diseases is still limited. The primary reason for this limitation is the unavailability of the crystal structure of the complex of PTP-LAR, and only a small number of

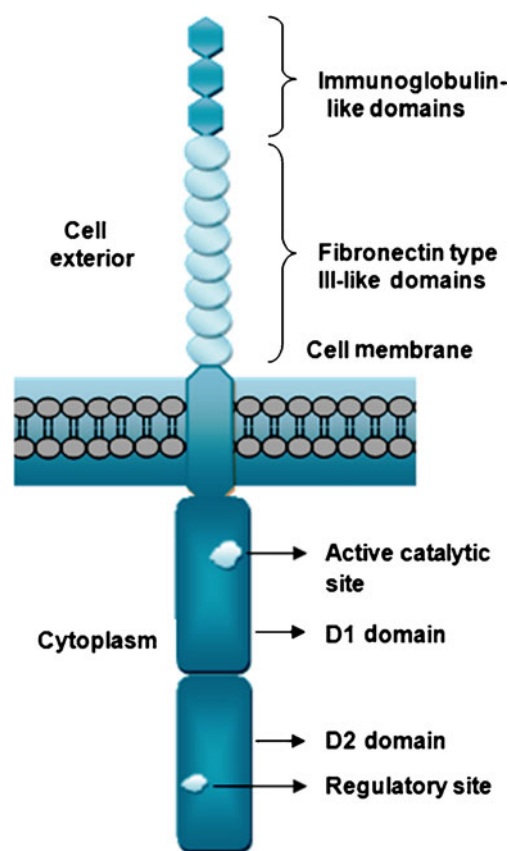


Fig. 1 Schematic representation of the structure of PTP-LAR

compounds are reported to be LAR inhibitors. Therefore, designing inhibitors for PTP-LAR is a considerable challenge. The purpose of the current study was to design and optimize potent LAR inhibitors and to validate the designed inhibitors using molecular docking studies. In the present study, a qualitative HipHop model was generated based on the ligand pharmacophore query. This HipHop model represents the 3D arrangement of the functional groups that are common to a set of molecules which interact with its biological target. A sequence of *in silico* strategies including pharmacophore mapping, database searching, SADMET screening and molecular docking was performed to identify the best compound. The toxicities of the compounds obtained were then predicted with the aid of Derek (Deductive Estimation of Risk from Existing Knowledge), in order to screen for safe and potent PTP-LAR inhibitors. [42–44].

Materials and methods

Data set

Molecules that are reported to be PTP-LAR inhibitors in the literature were collected and classified into two categories

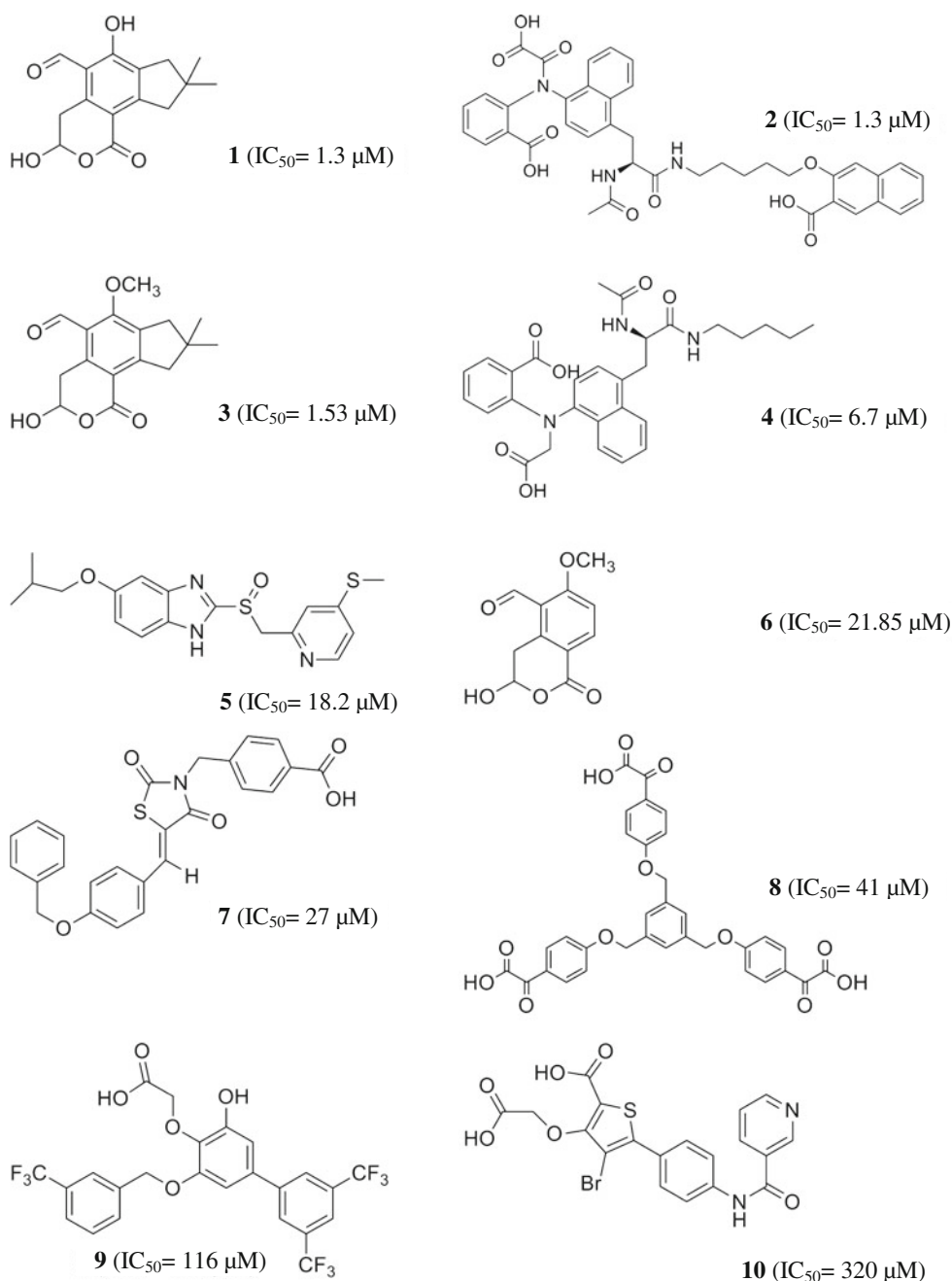
based on their IC_{50} values [45–48]. Molecules with IC_{50} values that are less than 10 μM were categorized as highly active molecules, and molecules with IC_{50} values between 10 μM and 350 μM were categorized as moderately active molecules [49–51]. Ten molecules (Chart 1) were sketched and minimized in Discovery Studio 2.5 using the Smart Minimization algorithm with a RMS gradient of 0.1 and the max step set at 200 [52]. In the Smart Minimization method, the molecules were first minimized with the steepest descent method. This was followed by the conjugate gradient method, which enables faster conver-

gence towards a local minimum. The HipHop pharmacophore models for the ten molecules were then generated.

Pharmacophore model generation

In pharmacophore modeling, a group of active and inactive molecules is analyzed to identify the key features of the pharmacophore present in the active molecules [53]. The chemical features of the resulting pharmacophore model reflect the characteristics of the atoms present in that molecule [54, 55]. In the present work, the common feature

Chart 1 The ten data set molecules used for pharmacophore generation, along with their experimental IC_{50} values



pharmacophore generation module in Discovery Studio 2.5 was used to develop different hypotheses. In this module, different conformations are generated for the given molecules, and these molecular conformations are superimposed on one another to generate a common pharmacophore feature for the overlapped portion. The feature mapping protocol was also used to examine the different features present in each molecule (Table 1). The five features selected to generate the pharmacophore models were hydrophobic (H), hydrogen-bond donor (D), aromatic ring (R), negative ionization (N) and hydrogen-bond acceptor (A) features. Using these two methods, we generated twenty different hypotheses from a data set of ten molecules [56, 57].

In method I, we generated pharmacophore models by assuming that all of the molecules have important features. The data set for the ten molecules was then used to generate the pharmacophore model. Method I generated ten different hypotheses with the help of both highly and moderately active molecules. The highly active molecules were specified by assigning a principal value of 2 and a MaxOmitFeat value of 0. The moderately active molecules were specified by assigning a principal value of 1 and a MaxOmitFeat value of 1. A principal value of 1 ensures that at least one mapping for each of the generated hypotheses will identify all of the chemical features of the molecules while generating the hypothesis. A MaxOmitFeat value of 1 ensures that all but one feature are mapped.

In method II, the models were generated by assuming that molecules with IC_{50} values of less than 10 μ M will generate the potent pharmacophore features necessary for the inhibition. Method II generated ten different hypotheses by considering only the highly active molecules (molecules 1, 2, 3, 4) from the dataset. The active molecules were specified by assigning them a principal value of 2 and a

MaxOmitFeat value of 0 (zero). A principal value of 2 ensures that all of the chemical features of the molecules are considered during hypothesis generation. A MaxOmitFeat value of 0 ensures that all of the features of all of the molecules are mapped. The seven common features selected for hypothesis generation were hydrophobic (H), hydrogen-bond donor (D), hydrogen-bond acceptor (A), ring aromatic (R), negative ionization (N), hydrophobic aliphatic (Z) and hydrophobic aromatic (Y) features.

Pharmacophore-based virtual screening

A pharmacophore-based database search approach was used to identify the ligands that map to the hypothesis [58–61]. The best pharmacophore model was selected and used to screen several 3D databases, including Specs, NCI, MiniMayBridge and InterBioScreen (IBS). The ligands bound with the protein phosphatases (PTPs) were extracted from the Protein Data Bank (PDB), and an in-house chemical database of 561 molecules was prepared [62, 63]. Overall, there were about 8,91,323 molecules in all five databases. The best and most flexible search method was applied to screen the database in order to retrieve putative hits. These molecules had their chemical moieties spatially mapped to their corresponding features in the pharmacophore query.

SADMET analysis

The hits obtained from the 3D database search were further subjected to SADMET filtration. The first filter applied was Lipinski's rule of five (molecular weight <500, HBD <5, HBA <10, and $\log P$ <5), and it screened molecules for their drug-like properties and oral bioavailability [64, 65]. The SADMET descriptors were then calculated by following the

Table 1 Pharmacophore chemical features for each molecule from the dataset

Sl. no.	Molecule	IC_{50} (μ M)	Principal	Max omit	Chemical features				
					Hydrogen bond acceptor (A)	Ring aromatic (R)	Hydrophobic feature (H)	Negative ionizable feature (N)	Hydrogen bond donor (D)
1	1	1.30	2	0	5	1	2	0	2
2	2	1.30	2	0	8	5	6	3	1
3	3	1.53	2	0	5	1	3	0	1
4	4	6.7	2	0	7	3	4	2	0
5	5	18.2	1	1	5	3	4	0	1
6	6	21.85	1	1	5	1	2	0	1
7	7	27	1	1	5	3	3	1	0
8	8	41	1	1	9	4	4	3	0
9	9	116	1	1	5	3	6	1	1
10	10	320	1	1	8	3	3	2	1

ADMET descriptors protocol in Discovery Studio 2.5. This protocol evaluates the QSAR models in order to estimate a range of ADMET-related properties for small molecules [66]. The descriptors selected for the calculation were aqueous solubility, percent human intestinal absorption (HIA), plasma protein binding (PPB), ADMET AlogP98, ADMET PSA 2D (fast polar surface area), cytochrome P450 (CYP450) 2D6 inhibition, and hepatotoxicity. Finally, pharmacophore max fitvalue was used as a third filter for virtual screening. This value indicates the quantitative pharmacophore features generated in a given hypothesis [57, 67].

Docking

The docking studies were carried out on the 1LAR crystal structure obtained from the Protein Data Bank (PDB) [68–70]. The protein structure was prepared for docking simulations using the Prepare Protein Wizard implemented in the Schrödinger package with a minimization RMSD of 0.3 Å by applying the OPLS_2005 force field. The Ligprep module was employed to prepare the ligands retrieved from the 3D database. They were then docked into the active site of the prepared protein using the GLIDE5.1 module from the Schrödinger suite [71]. An extra precision (XP) flexible GLIDE docking simulation was carried out by performing 1000 poses per docking run, yielding an output of five poses per ligand with an RMSD deviation of less than 0.5 Å. Glide uses a stochastic search docking algorithm to search for ligand positions and orientations through the Monte Carlo sampling of pose conformations in order to obtain an accurate docked pose. The Maestro 9.0 interface was used to visualize and analyze the active site and docking interactions. The best hits retrieved from GLIDE were then docked in 1LAR using the GOLD program (Cambridge Data Center, Cambridge, U.K.) [72]. For this, the protein was again prepared using Sybyl 7.1 (Tripos International, St. Louis, MO, USA) and minimized by applying the Amber 99 forcefield. The Gasteiger–Hückel charge was set with maximum number of iterations of 500 and a gradient of $0.05 \text{ kcal mol}^{-1} \text{ Å}^{-1}$ [73]. GOLD uses an evolutionary genetic algorithm to explore the full range of ligand conformational flexibility, and it allows for partial flexibility of the target protein. The docking simulation was carried out by performing 100,000 genetic algorithm operations, and it was terminated if the top three poses were within 1.5 Å RMSD.

Toxicity studies (Derek)

The screened molecules from the docking studies had their toxicities predicted using Deductive Estimation of Risk from Existing Knowledge (Derek) [74]. The Deductive

Estimation of Risk is a knowledge-based system that focuses on molecular substructures associated with qualitative and semiquantitative known toxicological endpoint prediction. The different toxicological endpoints selected for the prediction were carcinogenicity, chromosome damage, genotoxicity, hepatotoxicity, hERG channel inhibition, mutagenicity, ocular toxicity, reproductive toxicity, respiratory sensitization, skin sensitization, and thyroid toxicity [75]. These descriptor predictions are knowledge-based and several rule-based systems. Derek compares the different molecular structures and the alerts for these structures, which are responsible for the toxicity endpoints [76].

Results and discussion

Selection of the pharmacophore

In method I, the HipHop run generated ten hypotheses containing three hydrogen-bond acceptors (A), one ring aromatic (R) and one hydrophobic (H) feature with rank scores of between 91.175 and 93.721 (Fig. 2). The high rank scores obtained for all ten hypotheses indicated that these molecules were less likely to fit the chance correlation hypothesis (Table 2). Method I ensures that the minimum important features required for the molecules to act as PTP-LAR inhibitors are present. In method II, the HipHop run generated ten hypotheses containing three hydrogen-bond acceptors (A), one ring aromatic (R) and one hydrophobic aliphatic (Z) feature with rank scores of between 51.883 and 50.505. The difference between the rank scores of the

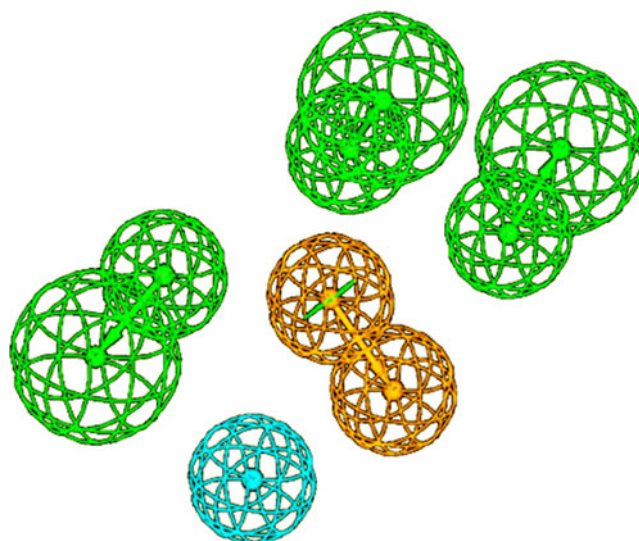


Fig. 2 Method I pharmacophore hypothesis (Hypo 1) (*green* three hydrogen bond acceptors, *cyan* one hydrophobic feature, *orange* one aromatic ring)

Table 2 Pharmacophore chemical features of the hypotheses generated using methods I and II

Sl. no.	HipHop hypotheses	Method I				Method II				Max fit
		Features	Ranking score	Direct hit	Partial hit	Features	Ranking score	Direct hit	Partial hit	
01	Hypo 1	RHAAA	93.721	11111111	00000000	RZAAA	51.883	1111	0000	5
02	Hypo 2	RHAAA	92.775	11111111	00000000	RZAAA	51.394	1111	0000	5
03	Hypo 3	RHAAA	92.715	11111111	00000000	RZAAA	51.394	1111	0000	5
04	Hypo 4	RHAAA	92.379	11111111	00000000	RZAAA	50.986	1111	0000	5
05	Hypo 5	RHAAA	92.326	11111111	00000000	RZAAA	50.986	1111	0000	5
06	Hypo 6	RHAAA	92.069	11111111	00000000	RZAAA	50.903	1111	0000	5
07	Hypo 7	RHAAA	91.541	11111111	00000000	RZAAA	50.651	1111	0000	5
08	Hypo 8	RHAAA	91.304	11111111	00000000	RZAAA	50.541	1111	0000	5
09	Hypo 9	RHAAA	91.224	11111111	00000000	RZAAA	50.541	1111	0000	5
10	Hypo 10	RHAAA	91.175	11111111	00000000	RZAAA	50.505	1111	0000	5

A hydrogen-bond acceptor, *H* hydrophobic, *R* ring aromatic, *Z* aliphatic hydrophobic

different hypotheses is very small (1.378), and all ten hypotheses picked out the same features (Table 2). The hypotheses generated in methods I and II indicate the essential features needed for the inhibition of PTP-LAR. Out of the five features generated by both methods, four features were picked out by both methods: three hydrogen-bond acceptors (*A*) and one ring aromatic (*R*). The feature that was picked out by method I but not II was the hydrophobic (*H*) feature, while that picked out by method II but not I was the aliphatic feature that is hydrophobic (*Z*) in nature. Since method II was generated using highly active molecules ($IC_{50} < 10 \mu M$), the general hydrophobic feature picked up by method I was elucidated more precisely by method II to be an aliphatic hydrophobic and not an aromatic hydrophobic feature. Thus, the hypothesis generated in method II was considered for further screening. Both Hypo 2 and Hypo 3 have the same ranking scores, which suggests that they are statistically identical and that the molecules have the same probability of fitting the chance correlation hypothesis. When Hypo 2 and Hypo 3 were merged together, all of the features overlapped exactly

except for the ring aromatic (*R*) feature (Fig. 3a). Hypo 2 of method II mapped all of the features in the active molecules and showed good correlations among the best fit value, relative energy and activity. In addition, the aromatic feature (*R*) picked up in Hypo 2 was oriented towards an active residue, Tyr1355, thus increasing the probability of π - π interactions, which is essential for inhibition. Therefore, Hypo 2 lacked any bias and it was selected as the best hypothesis for the virtual screening of different databases (Fig. 3b). Hypo 2, with a rank score of 51.394 and a max fit value of 5, satisfied the 3D arrangement of all of the chemical features present in highly active molecule 2 from the dataset shown in Fig. 3c. The pharmacophore model was then validated via docking interactions with the standard molecules of the data set.

Pharmacophore-based virtual screening

The best hypothesis, Hypo 2, selected using method II was used to screen the 3D molecular databases. The first screenings were carried out with the NCI, MiniMaybridge,

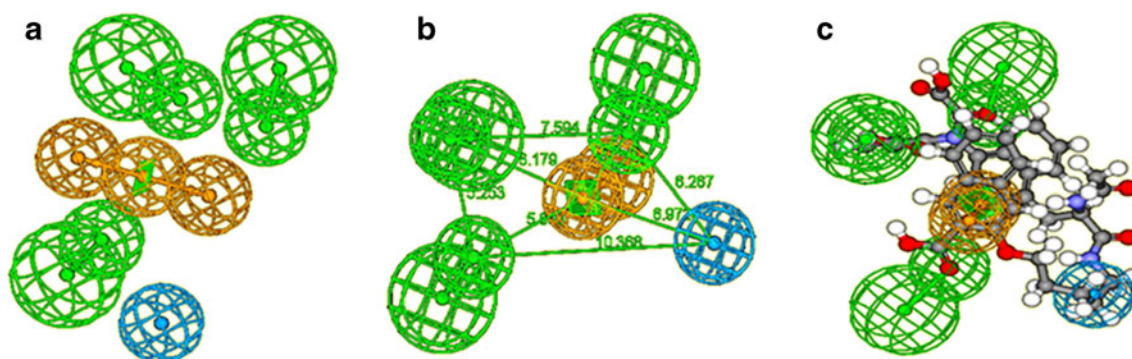


Fig. 3 a–c Method II pharmacophore hypothesis: **a** superimposition of Hypo 2 and Hypo 3; **b** best selected Hypo 2 with distance constraints; **c** mapping of highly active molecule 2 onto Hypo 2

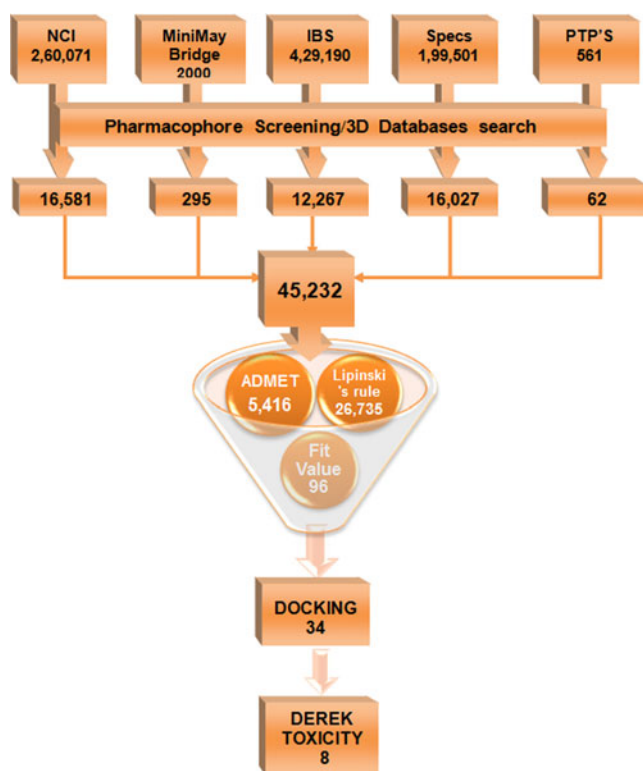
Interbioscreen (IBS), Specs, and in-house PTP databases, and yielded 16,581, 295, 12,267, 16,027, and 62 molecules, respectively. A total of 45,232 molecules were retrieved and subjected to further virtual screening, as shown in the flow chart in Scheme 1.

SADMET analysis

The SADMET analysis resolved a common problem encountered when designing new PTP inhibitors: poor bioavailability [77, 78]. The first filter used was Lipinski's rule of five, and this retrieved a total of 26,735 molecules. The second filter used was SADMET properties. Based on the descriptor cutoff values mentioned in Table 3, the molecules were assigned to good, moderate, poor, and very poor prediction levels. Only molecules showing good and moderate prediction levels were then employed to retrieve 5416 drug-like molecules. Finally, by considering a pharmacophore maxfit value of 3 to be a filter, 96 molecules were obtained, and these were further validated by docking studies, as shown in Scheme 1.

Docking analysis

Since the crystal structure of PTP-LAR was available, the screened molecules were further validated by performing



Scheme 1 Schematic representation of the sequence of in silico strategies used for virtual screening followed by molecular docking and Derek toxicity prediction

docking studies. The 96 molecules obtained from the SADMET screening were docked into the active site using GLIDE, and this yielded several poses for each ligand. Scores were then assigned based on their Glide energies, docking gscores, E-model energies, and top-ranked poses. The top 100 poses of the Glide gscore (cutoff value: -4.68) and those exhibiting the greatest hydrogen contact (cutoff value: 4) were screened to retrieve 34 molecules. These 34 molecules, in addition to another four highly active molecules ($IC_{50} < 10 \mu\text{M}$) from the dataset, were further validated in GOLD docking studies. The GOLD score fitness function along with the descriptor contact count were considered in order to predict the best ligand binding position (Table 4). The docking simulation carried out showed that the molecules buried into the active-site pocket exhibited better interactions and GOLD fitness values than the known inhibitor illudalic acid (molecule 1: $IC_{50} < 1.3 \mu\text{M}$). The docking poses of the screened 34 molecules were superimposed onto the pharmacophore model Hypo 2 of method II. The results indicated that 29 of the 34 molecules mapped onto the pharmacophore. The mapping of the best screened molecule from each database is shown in Fig. 4. This demonstrated that the interactions indicated in the docking studies were indeed crucial ligand–protein interactions. All of the screened ligands showed interactions with the active residues (Tyr1355, Asp1490, Ser1523, and Gln1566) that are important for achieving strong binding affinity. These interactions involved both strong hydrogen-bonding, π – π , and π –cation interactions that are essential for stabilizing the ligand in the active site cavity. In addition, the ligands also showed van der Waals, hydrophobic and apolar interactions with other residues of the active binding site (Asn1428, His1521, Arg1528, and Thr1529). The docking analysis revealed that bicyclic ring compounds had the highest rankings with high GOLD fitness scores, indicating the importance of the bicyclic scaffold to LAR inhibition. Also, the docking scores improved when a sulfonamide group was present in the inhibitors. Further analysis revealed that inhibitors which have carboxylic acid, amide, catechol, or ketone ring structures in their side chains exhibited more hydrogen-bonding interactions with the protein. Detailed discussions of the best screened molecule from each database are provided below (see also Fig. 5). To allow a clearer understanding of the docking interaction, data in the form of 2D images showing the bond distances, scores, atom numbers, etc. are provided in the “[Electronic supplementary material](#).”

Illudalic acid (molecule 1)

The docking simulation of illudalic acid on PTP-LAR (Fig. 5a) showed that the hydroxyl group at the 3 position

Table 3 Cutoff values used to assign the prediction levels of various SADMET descriptors

Properties	SADMET descriptors	
	Good prediction levels	Moderate prediction levels
Solubility	ADMET solubility (−4 to −2)	ADMET solubility (−2 to 0)
Absorption	ADMET absorption level 0 (less than 95%)	ADMET absorption level 1 (less than 99%)
Distribution	ADMET <i>AlogP</i> 98 4.0 to 5.0 (>95%)	ADMET <i>AlogP</i> 98 <4.0 (<90%)
Metabolism	ADMET CYP2D6 probability<0.5 (non-inhibitors)	
Toxicity	ADMET hepatotoxicity probability<0.5 (non-toxic)	

interacts with the active residue Gln1566 through a strong hydrogen bond (2.26 Å). The oxygen of the 5-carbaldehyde group showed strong hydrogen bonding (2.96 Å) with Arg1528, as reported by Qing Ling et al. [48]. Apart from

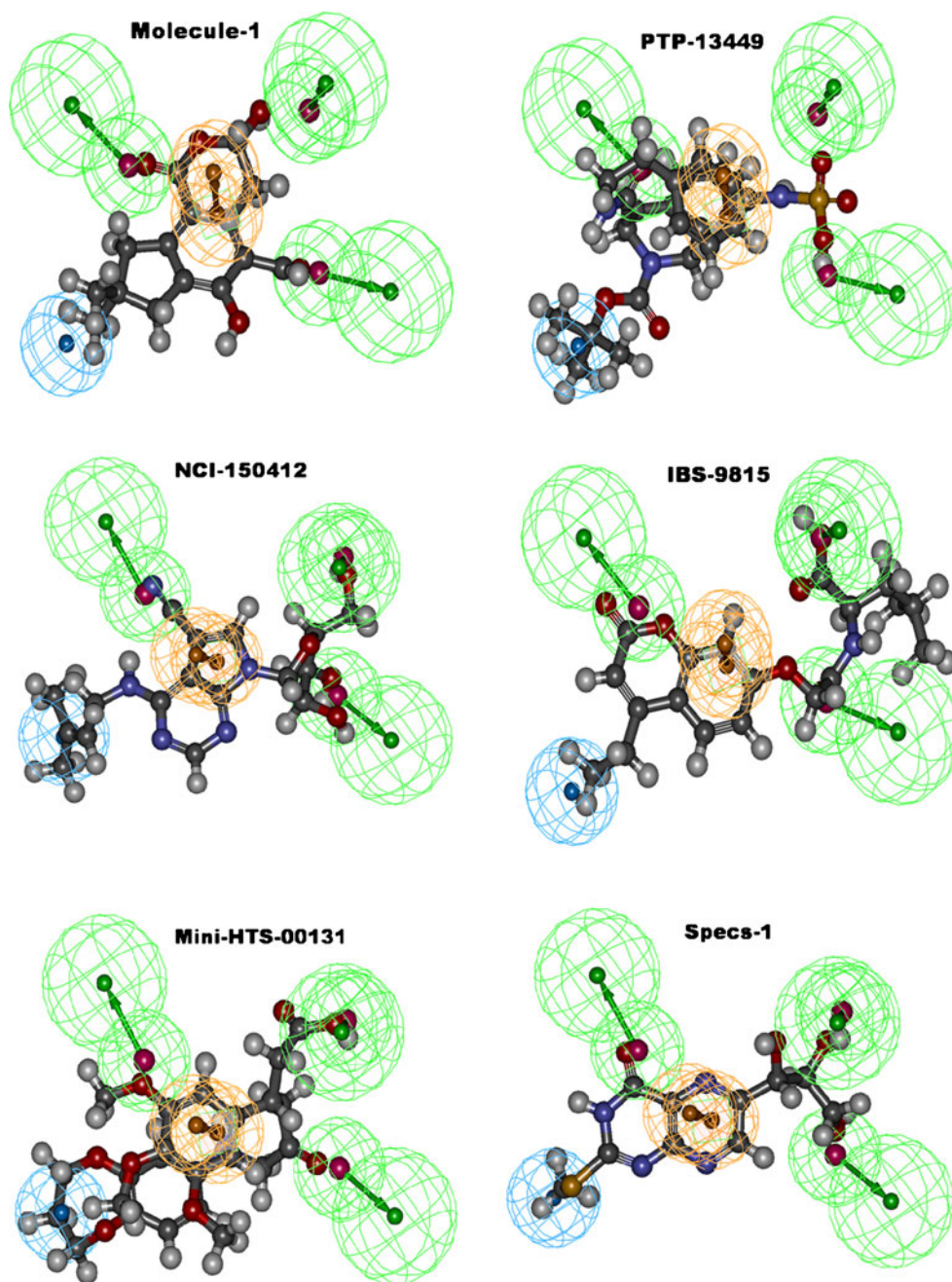
the above interactions, the oxygen of the 1-oxo group formed a hydrogen bond (2.59 Å) with Lys1433, and the phenyl ring also showed a π -cation interaction with Arg1528. The ligand orientation analysis performed in this

Table 4 Docking results for screened inhibitors with the GOLD fitness score for the best dock pose in each case

Sl. no.	Entry	Descriptor contact count ^a	GOLD fitness score
1	NCI-150412 dock1	15	53.92
2	PTP-13449 dock2	12	54.54
3	NCI-295273 dock1	12	54.36
4	Specs 1 dock2	11	48.73
5	IBS-9815 dock4	11	58.52
6	NCI-694770 dock3	11	46.83
7	Molecule 2 dock2	11	41.56
8	Specs 2 dock4	10	28.29
9	IBS-20311 dock3	10	43.93
10	IBS-41527 dock10	10	52.58
11	IBS-41908 dock4	10	70.58
12	Mini-HTS 00131 dock3	10	55.02
13	NCI-121926 dock2	10	50.65
14	Molecule 4 dock9	10	35.33
15	Specs 3 dock10	9	56.42
16	PTP-14240 dock3	9	47.66
17	PTP-14241 dock3	9	46.97
18	IBS-13297 dock2	9	64.86
19	IBS-17751 dock7	9	43.50
20	NCI-114346 dock2	9	47.54
21	NCI-232956 dock3	9	45.85
22	NCI-628338 dock3	9	39.35
23	NCI-694770 dock1	9	46.16
24	Molecule 1 dock2	9	38.96
25	Molecule 3 dock1	8	34.77
26	IBS-20826 dock5	8	55.13
27	IBS-21906 dock1	8	49.89
28	Specs 4 dock1	8	54.75
29	Mini-GK 03286 dock2	8	62.28
30	Mini-RJC 03319 dock7	8	49.02

^a Descriptor contact count is the calculated maximum number of possible contacts a molecule can make with the receptor (via hydrogen bonds, π -interactions, van der Waals and hydrophobic interactions)

Fig. 4 a–f Mapping of the best screened lead molecule from each database onto the pharmacophore model Hypo 2 of method II: **a** molecule 1 (illudalic acid) from the data set; **b** PTP-13499; **c** NCI-150412; **d** IBS-9815; **e** HTS-00131; **f** Specs 1



study was very similar to the analysis performed by Qing Ling et al.

PTP-13449

The docking analysis of PTP-13449 (Fig. 5b) showed the highest number of hydrogen-bonding interactions and the second highest descriptor contact count with the active site (Table 4). The tetrahydroisoquinoline ring buried in the active pocket showed favorable hydrophobic interactions. The carbamoyl group attached at the 3 position formed two strong hydrogen bonds (2.50 Å and 2.74 Å) with the

active residues Tyr1355 and Ser1523. These residues are important in dephosphorylation. The sulfamic acid attached at the 7-position on the tetrahydroisoquinoline ring played a major role in the binding by providing the greatest number of hydrogen-bonding interactions. It showed four hydrogen-bond acceptor interactions with Arg1528 (3.1 Å), Trp1488 (2.74 Å), and Gln1570 (2.67 Å, 3.08 Å). It also showed two backbone hydrogen-bond donor interactions with the Gly1492 (2.5 Å, 2.7 Å) residue. The oxygen of *tert*-butoxy at the 2 position formed a hydrogen bond with the exposed active residue Gln1566 (3.07 Å). The phenyl ring attached to the side

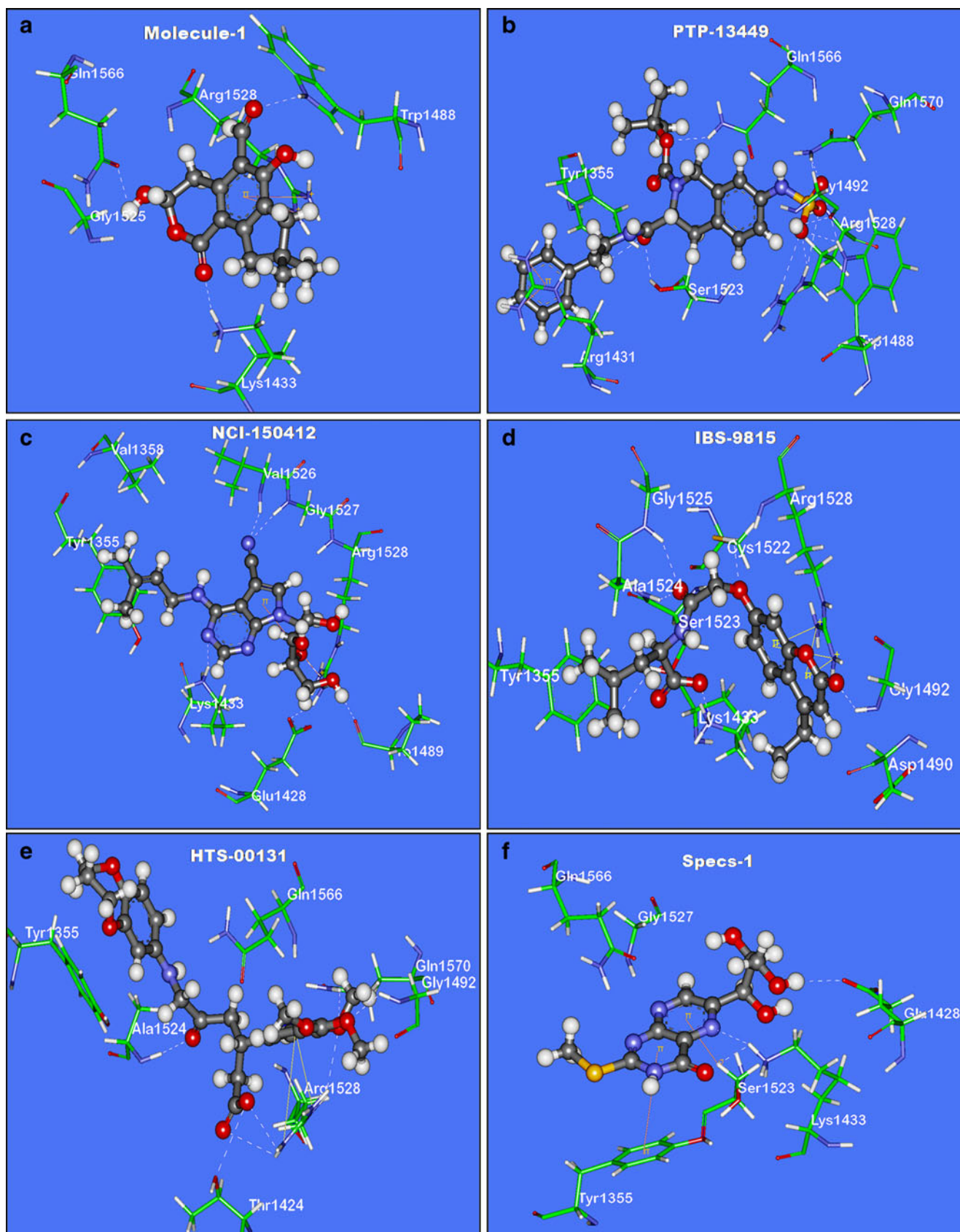


Fig. 5 a–f Three-dimensional GOLD docking pose models of the best screened lead molecule from each database: **a** molecule 1 (illudalic acid) from the data set; **b** PTP-13499; **c** NCI-150412; **d** IBS-9815; **e** HTS-00131; **f** Specs 1, all bound to the PTP-LAR active site. Hydrogen bonds are indicated by *white lines*, π - π and π -cation interactions are indicated by *orange lines*

chain at the 2 position also showed two π -cation interactions with Arg1431.

NCI-150412

The docking of NCI-150412 at the active site showed the highest descriptor contact count (Fig. 5c). In addition to the π -cation interaction with Lys1433, the pyrrole ring that docked in the center of the cavity also showed a hydrophobic interaction. The 5-carbonitrile group showed strong hydrogen-bonding interactions with Val1526 (3.13 Å) and Gly1527 (3.18 Å). The oxygen atoms from the tetrahydrofuran ring and the hydroxymethyl formed strong hydrogen bonds with Arg1528 (2.53 Å, 2.9 Å), an active site residue. The hydroxyl group at the 5 position in the hydroxymethyl and at the 3 position on the tetrahydrofuran ring showed strong hydrogen-bonding interactions with Pro1489 (1.69 Å) and Gly1492 (3.01 Å, 3.05). Also, the exposed side chain of the 4-(3-methylbut-2-enylamino)

group formed a hydrogen bond with the active residue Tyr1355.

IBS-9815

The IBS compound 9815 was selected as the best compound from the docking studies (Fig. 5d). The oxygen at the 1 position of the pentanoic acid forms two strong hydrogen-bonding interactions with the active residues Ser1523 (2.64 Å) and Tyr1355 (2.71 Å). The anionic oxygen (O^-) group of the same acid also participates in a hydrogen bond with Lys1433 (2.75 Å). The phenyl ring exhibited a strong π -cation interaction with the active residue Arg1528, and the oxo group attached to the ring formed a hydrogen bond with Gly1492 (2.36 Å). Additionally, a hydrogen-bond acceptor interaction was observed between the oxygen group at the 4 position of the pentanoic acid and Ala1524 (2.35 Å).

Mini-HTS 00131

HTS 00131 was the best hit obtained from the MiniMaybridge database. Three hydrogen bonds and a π -cation bond were formed between the receptor and ligand (Fig. 5e). The carbonyl oxygen group of hexanoic acid in

Table 5 Derek toxicity predictions showing toxic endpoints and structural alerts

Sl. no.	Compound name	Endpoint	Alert name
1	NCI-150412	Nothing	
2	NCI-295273	Carcinogenicity	Substituted pyrimidine or purine
3	PTP-13449	Nothing	
4	Specs 1	Nothing	
5	NCI-121926	Carcinogenicity	Substituted pyrimidine or purine
6	IBS-9815	Peroxisome proliferation Skin sensitization Photoallergenicity	β -O/S-substituted carboxylic acid or precursor Phenyl ester resorcinol or precursor Coumarin derivatives
7	NCI-694770	Nothing	
8	IBS-41908	Peroxisome proliferation	β -O/S-substituted carboxylic acid or precursor
9	Specs 2	Thyroid toxicity Phototoxicity Skin sensitization	4-Aminoaryl sulfonamide or precursor Aryl sulfonamide Phenol or precursor
10	Mini-HTS 00131	Nothing	
11	IBS-41527	Peroxisome proliferation	β -O/S-substituted carboxylic acid or precursor
12	NCI-232956	Nothing	
13	Molecules 1, 3 and 6	Mutagenicity Genotoxicity Chromosome damage Skin sensitization	Alkyl aldehyde or precursor, aldehyde precursor
14	Molecules 2, 4, 5, 7 and 8	Nothing	
15	Molecules 9 and 10	Peroxisome proliferation	β -O/S-substituted carboxylic acid or precursor

the side chain showed a strong hydrogen-bonding interaction with Thr1424 (2.55 Å). The oxygen of the 4-methoxy group in the 3-trimethoxyphenyl ring showed an interaction with Gly1492 through a strong hydrogen bond (2.82 Å). The 3-trimethoxyphenyl ring also participated in a strong π -cation interaction with Arg1528 at the active site. The 5-oxo group of hexanoic acid formed a hydrogen bond with Ala1524 (2.75 Å). Among the ligands, this ligand also exposed the largest portion to the receptor and offered good ligand–receptor interactions.

Specs 1

The compound selected for the docking simulation from the Specs database followed the same binding pattern as NCI-694770 (Fig. 5f). Though this compound showed a few

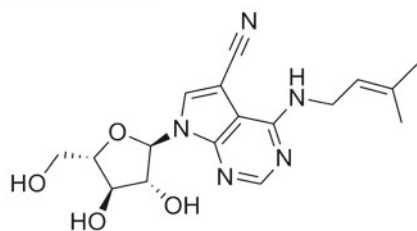
hydrogen-bonding interactions, its docked pose was similar to the orientation of the reported illudalic acid. The keto group at the 4 position of pteridinone in the side chain formed strong hydrogen bonds with two residues: Tyr1355 (2.87 Å) and Lys1433 (2.77 Å). The nitrogen at the 5 position on the pteridinone ring was also a strong hydrogen-bond acceptor for Lys1433 (3.31 Å).

Toxicity studies (Derek)

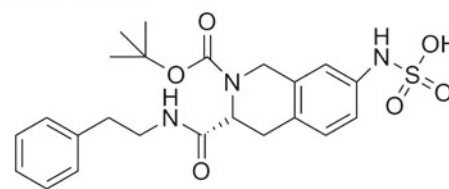
Toxicity testing with the help of Derek increases the chance of selecting the most promising molecules (Table 5). In general, many of the compounds showed skin sensitization because of the presence of toxic moieties such as phenol, resorcinol and 1,3-diketone groups. Compounds with coumarin and aryl sulfonamide groups also showed

Chart 2 2D chemical structures of the final eight selected hits from the virtual screening procedure

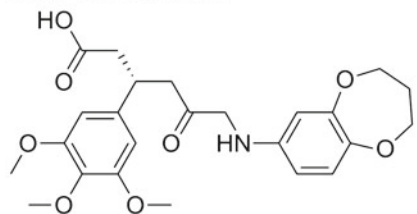
NCI-150412



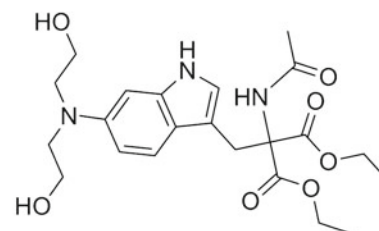
PTP-13449



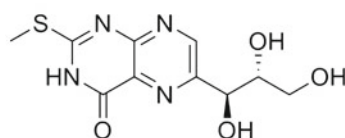
Mini-HTS 00131



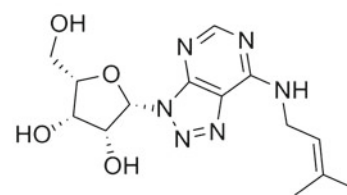
NCI-114346



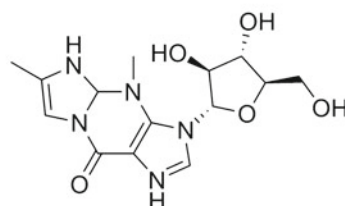
Specs-1



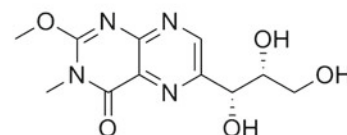
NCI-232956



NCI- 628338



NCI- 694770



phototoxicity. The majority of the compounds from the IBS database, such as IBS-9815, IBS-41908 and IBS-41527, were predicted to be peroxisome proliferators due to the presence of β -O/S-substituted carboxylic acids or their precursors. The compounds with an alkyl aldehyde group or its precursor, along with substituted vinyl ketone, were predicted to show genotoxicity and chromosome damage. NCI compounds with substituted pyrimidine or purine groups, such as NCI-295273 and NCI-121926, exhibited carcinogenicity. The compound Specs 2 from the Specs database showed thyroid toxicity due to the presence of 4-aminoaryl sulfonamide or its precursor. The known inhibitors molecules 1, 3 and 6 were predicted to exhibit mutagenicity, genotoxicity, chromosome damage and skin sensitization, whereas molecules 9 and 10 showed peroxisome proliferation. Out of the 34 ligands docked, eight were predicted to be safe and free from toxicological endpoints. These eight compounds were obtained from the four databases used in the screening studies: one compound each from the in-house PTP database (PTP-13449), Mini-MayBridge (HTS 00131), the Specs database (Specs 1), and five from NCI (NCI-150412, NCI-114346, NCI-232956, NCI- 628338 and NCI- 694770) (see Chart 2).

Conclusions

We have successfully identified new potent inhibitors using a unique sequence of *in silico* strategies that involved pharmacophore studies, SADMET-based virtual screening and docking followed by Derek toxicity studies. This method helped us to screen and select eight compounds with different scaffolds that show high binding affinity for PTP-LAR from databases containing 8,91,323 compounds. The pharmacophore studies carried out in the present study are the first contributions toward the rational design of PTP-LAR inhibitors. The best pharmacophore hypothesis consists of three hydrogen-bond acceptor (A), one ring aromatic (R), and one hydrophobic aliphatic (Z) feature. In order to overcome the common bioavailability problem associated with phosphatases, the hit compounds were filtered using Lipinski's rule of five and evaluated for their SADMET properties. This retrieved compounds that showed drug-like properties and good orally bioavailability, fulfilling the *in silico* cutoff criteria. The results obtained from the docking studies suggested not only the essential ligand binding interactions but also the ligand binding patterns necessary for LAR inhibition. The docking analysis showed that compounds with isoquinoline or naphthalene as a basic scaffold and a methoxycarbonyl side chain can be assumed to act as potent PTP-LAR inhibitors. The results obtained from the docking studies showed better hydrogen-bonding interactions and docking

fitness scores than previously known inhibitors. The ligand pharmacophore mapping studies further validated the screened protocol and confirmed that the final screened molecules can be presumed to show potent inhibitor activity. The docking analysis also made it clear that compounds with a sulfonamide group gave the best docking scores and showed more favorable interactions, indicating that sulfonylurea derivatives may be employed as phosphatase drugs in the future. The 3D information provided by these docking and pharmacophore mapping studies were meaningful, as they indicated the interactions required for LAR inhibition. The molecular series PTP-13400 and PTP-13900 from the PTP database showed favorable GOLD fitness scores and could be further explored in *in vitro* studies. The Derek toxicity studies revealed eight screened compounds that were safe and free from toxic endpoints. The compounds PTP-13449 and NCI-150412 should be further explored by experimental analysis in order to obtain potential lead compounds for the treatment of obesity and diabetes disease.

Acknowledgments The authors thank the Department of Pharmaceuticals, the Minister of Chemicals and Fertilizers, the Department of Science and Technology (DST), and the Council of Scientific and Industrial Research (CSIR) New Delhi for financial assistance.

References

1. Stoker AW (2005) Protein tyrosine phosphatases and signalling. *J Endocrinol* 185:19–33
2. Stone RL, Dixon JE (1994) Protein-tyrosine phosphatases. *J Biol Chem* 269:31323–31326
3. Ventura JJ, Nebreda Á (2006) Protein kinases and phosphatases as therapeutic targets in cancer. *Clin Transl Oncol* 8:153–160
4. Hunter T (2000) Signaling—2000 and beyond. *Cell* 100:113–127
5. Kappert K, Peters KG, Bohmer FD, Ostman A (2005) Tyrosine phosphatases in vessel wall signaling. *Cardiovasc Res* 65:587–598
6. Stoker AW (2001) Receptor tyrosine phosphatases in axon growth and guidance. *Curr Opin Neurobiol* 11:95–102
7. Zhang ZY (2005) Functional studies of protein tyrosine phosphatases with chemical approaches. *Biochim Biophys Acta* 1754:100–107
8. Andersen JN, Mortensen OH, Peters GH, Drake PG, Iversen LF, Olsen OH, Jansen PG, Andersen HS, Tonks NK, Moller NPH (2001) Structural and evolutionary relationships among protein tyrosine phosphatase domains. *Mol Cell Biol* 21:7117–7136
9. Tonks NK (2006) Protein tyrosine phosphatases: from genes, to function, to disease. *Nat Rev Mol Cell Biol* 7:833–846
10. Zhang ZY, Wang Y, Wu L, Fauman EB, Stuckey JA, Schubert HL, Saper MA, Dixon JE (1994) The Cys (X) 5Arg catalytic motif in phosphoester hydrolysis. *Biochemistry* 33:15266–15270
11. Tabernero L, Aricescu AR, Jones EY, Szedlacsek SE (2008) Protein tyrosine phosphatases: structure–function relationships. *FEBS J* 275:867–882
12. Tiganis T, Bennett AM (2007) Protein tyrosine phosphatase function: the substrate perspective. *Biochem J* 402:1–15

13. Alonso A, Sasin J, Bottini N, Friedberg I, Osterman A, Godzik A, Hunter T, Dixon J, Mustelin T (2004) Protein tyrosine phosphatases in the human genome. *Cell* 117:699–711
14. Bourdeau A, Dubé N, Tremblay ML (2005) Cytoplasmic protein tyrosine phosphatases, regulation and function: the roles of PTP1B and TC-PTP. *Curr Opin Cell Biol* 17:203–209
15. Paul S, Lombroso PJ (2003) Receptor and nonreceptor protein tyrosine phosphatases in the nervous system. *Cell Mol Life Sci* 60:2465–2482
16. Burke TR, Zhang ZY (1998) Protein-tyrosine phosphatases: structure, mechanism, and inhibitor discovery. *Biopolymers* 47:225–241
17. Zhang ZY (2003) Chemical and mechanistic approaches to the study of protein tyrosine phosphatases. *Acc Chem Res* 36:385–392
18. Asante-Appiah E, Kennedy BP (2003) Protein tyrosine phosphatases: the quest for negative regulators of insulin action. *Am J Physiol Endocrinol Metab* 284:663–670
19. Beltran PJ, Bixby JL (2003) Receptor protein tyrosine phosphatases as mediators of cellular adhesion. *Front Biosci* 8:87–99
20. Elchebly M, Cheng A, Tremblay ML (2000) Modulation of insulin signaling by protein tyrosine phosphatases. *J Mol Med* 78:473–482
21. Johnson KG, Van Vactor D (2003) Receptor protein tyrosine phosphatases in nervous system development. *Physiol Rev* 83:1–24
22. Kulas DT, Goldstein BJ, Mooney RA (1996) The transmembrane protein-tyrosine phosphatase LAR modulates signaling by multiple receptor tyrosine kinases. *J Biol Chem* 271:748–754
23. Nam HJ, Poy F, Krueger NX, Saito H, Frederick CA (1999) Crystal structure of the tandem phosphatase domains of RPTP LAR. *Cell* 97:449–458
24. Tsujikawa K, Kawakami N, Uchino Y, Ichijo T, Furukawa T, Saito H, Yamamoto H (2001) Distinct functions of the two protein tyrosine phosphatase domains of LAR (leukocyte common antigen-related) on tyrosine dephosphorylation of insulin receptor. *Mol Endocrinol* 15:271–280
25. Yang T, Bernabeu R, Xie Y, Zhang JS, Massa SM, Rempel HC, Longo FM (2003) Leukocyte antigen-related protein tyrosine phosphatase receptor: a small ectodomain isoform functions as a homophilic ligand and promotes neurite outgrowth. *J Neurosci* 23:3353–3363
26. Chagnon MJ, Uetani N, Tremblay ML (2004) Functional significance of the LAR receptor protein tyrosine phosphatase family in development and diseases. *Biochem Cell Biol* 82:664–675
27. Goldstein BJ (2002) Protein-tyrosine phosphatases: emerging targets for therapeutic intervention in type 2 diabetes and related states of insulin resistance. *J Clin Endocrinol Metab* 87:2474–2480
28. Kulas DT, Zhang WR, Goldstein BJ, Furlanetto RW, Mooney RA (1995) Insulin receptor signaling is augmented by antisense inhibition of the protein tyrosine phosphatase LAR. *J Biol Chem* 270:2435–2438
29. Li L, Dixon JE (2000) Form, function, and regulation of protein tyrosine phosphatases and their involvement in human diseases. *Semin Immunol* 12:75–84
30. Aicher B, Lerch MM, Muller T, Schilling J, Ullrich A (1997) Cellular redistribution of protein tyrosine phosphatases LAR and PTPsigma by inducible proteolytic processing. *J Cell Biol* 138:681–696
31. Pulido R, Serra-Pages C, Tang M, Streuli M (1995) The LAR/PTP delta/PTP sigma subfamily of transmembrane protein-tyrosine-phosphatases: multiple human LAR, PTP delta, and PTP sigma isoforms are expressed in a tissue-specific manner and associate with the LAR-interacting protein LIP. 1. *Proc Natl Acad Sci USA* 92:11686–11690
32. Ren JM, Li PM, Zhang WR, Sweet LJ, Cline G, Shulman GI, Livingston JN, Goldstein BJ (1998) Transgenic mice deficient in the LAR protein-tyrosine phosphatase exhibit profound defects in glucose homeostasis. *Diabetes* 47:493–497
33. Ahmad F, Considine RV, Goldstein BJ (1995) Increased abundance of the receptor-type protein-tyrosine phosphatase LAR accounts for the elevated insulin receptor dephosphorylating activity in adipose tissue of obese human subjects. *J Clin Invest* 95:2806–2812
34. Ahmad F, Goldstein BJ (1997) Functional association between the insulin receptor and the transmembrane protein-tyrosine phosphatase LAR in intact cells. *J Biol Chem* 272:448–457
35. Hashimoto N, Feener EP, Zhang WR, Goldstein BJ (1992) Insulin receptor protein-tyrosine phosphatases. *J Biol Chem* 267:13811–13814
36. Zabolotny JM, Kim YB, Peroni OD, Kim JK, Pani MA, Boss O, Klamann LD, Kamatkar S, Shulman GI, Kahn BB (2001) Overexpression of the LAR (leukocyte antigen-related) protein-tyrosine phosphatase in muscle causes insulin resistance. *Proc Natl Acad Sci USA* 98:5187–5192
37. Zhang WR, Li PM, Oswald MA, Goldstein BJ (1996) Modulation of insulin signal transduction by eutopic overexpression of the receptor-type protein-tyrosine phosphatase LAR. *Mol Endocrinol* 10:575–584
38. Miscio G, Tassi V, Coco A, Soccio T, Di Paola R, Prudente S, Baratta R, Frittitta L, Ludovico O, Padovano L (2004) The allelic variant of LAR gene promoter –127 bp T→A is associated with reduced risk of obesity and other features related to insulin resistance. *J Mol Med* 82:459–466
39. Phung TL, Mooney RA, Kulas DT, Sparks CE, Sparks JD (1997) Suppression of the protein tyrosine phosphatase LAR reduces apolipoprotein B secretion by McA-RH7777 rat hepatoma cells. *Biochem Biophys Res Commun* 237:367–371
40. Tautz L, Pellecchia M, Mustelin T (2006) Targeting the PTPome in human disease. *Expert Opin Ther Targets* 10:157–177
41. Zhang ZY (2001) Protein tyrosine phosphatases: prospects for therapeutics. *Curr Opin Chem Biol* 5:416–423
42. Good AC, Cheney DL, Sitkoff DF, Tokarski JS, Stouch TR, Bassolino DA, Krystek SR, Li Y, Mason JS, Perkins TDJ (2003) Analysis and optimization of structure-based virtual screening protocols. 2. Examination of docked ligand orientation sampling methodology: mapping a pharmacophore for success. *J Mol Graph Model* 22:31–40
43. Kapetanovic IM (2008) Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. *Chem Biol Interact* 171:165–176
44. McInnes C (2007) Virtual screening strategies in drug discovery. *Curr Opin Chem Biol* 11:494–502
45. Bialy L, Waldmann H (2005) Inhibitors of protein tyrosine phosphatases: next generation drugs. *Angew Chem Int Ed* 44:3814–3839
46. Dewang PM, Hsu NM, Peng SZ, Li WR (2005) Protein tyrosine phosphatases and their inhibitors. *Curr Med Chem* 12:1–22
47. Hoffman HE, Blair ER, Johndrow JE, Bishop AC (2005) Allele-specific inhibitors of protein tyrosine phosphatases. *J Am Chem Soc* 127:2824–2825
48. Ling Q, Huang Y, Zhou Y, Cai Z, Xiong B, Zhang Y, Ma L, Wang X, Li X, Li J (2008) Illudalic acid as a potential LAR inhibitor: synthesis, SAR, and preliminary studies on the mechanism of action. *Bioorg Med Chem* 16:7399–7409
49. McCain DF, Wu L, Nickel P, Kassack MU, Kreimeyer A, Gagliardi A, Collins DC, Zhang ZY (2004) Suramin derivatives as inhibitors and activators of protein-tyrosine phosphatases. *J Biol Chem* 279:14713–14725
50. Srinivasan R, Uttamchandani M, Yao SQ (2006) Rapid assembly and in situ screening of bidentate inhibitors of protein tyrosine phosphatases. *Org Lett* 8:713–716

51. Vintonyak VV, Antonchick AP, Rauh D, Waldmann H (2009) The therapeutic potential of phosphatase inhibitors. *Curr Opin Chem Biol* 13:272–283
52. Accelrys Inc. (2008) Discovery Studio 2.1. Accelrys Inc., San Diego
53. Oranit D, Alexandra SP, Ruth N (2004) Predicting molecular interactions in silico. I. A guide to pharmacophore identification and its applications to drug design. *Curr Med Chem* 11:71–90
54. Li HF, Lu T, Zhu T, Jiang YJ, Rao SS, Hu LY, Xin BT, Chen YD (2009) Virtual screening for Raf-1 kinase inhibitors based on pharmacophore model of substituted ureas. *Eur J Med Chem* 44:1240–1249
55. Sakkiiah S, Krishnamoorthy N, Gajendrarao P, Thangapandian S, Lee Y, Kim S, Suh JK, Kim HH, Lee KW (2009) Pharmacophore mapping and virtual screening for SIRT1 activators. *Bull Korean Chem Soc* 30:1153–1156
56. Doddareddy MR, Jung HK, Lee JY, Lee YS, Cho YS, Koh HY, Pae AN (2004) First pharmacophoric hypothesis for T-type calcium channel blockers. *Bioorg Med Chem* 12:1605–1611
57. Vadivelan S, Sinha BN, Rambabu G, Boppana K, Jagarlapudi S (2008) Pharmacophore modeling and virtual screening studies to design some potential histone deacetylase inhibitors as new leads. *J Mol Graph Model* 26:935–946
58. Greene J, Kahn S, Savoj H, Sprague P, Teig S (1994) Chemical function queries for 3D database search. *J Chem Inf Comput Sci* 34:1297–1308
59. Klebe G (2006) Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today* 11:580–594
60. Martin YC (1992) 3D database searching in drug design. *J Med Chem* 35:2145–2154
61. Walters WP, Stahl MT, Murcko MA (1998) Virtual screening—an overview. *Drug Discov Today* 3:160–178
62. Herman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242
63. Voigt JH, Bienfait B, Wang S, Nicklaus MC (2001) Comparison of the NCI open database with seven large chemical structural databases. *J Chem Inf Comput Sci* 41:702–712
64. Lipinski CA (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 44:235–249
65. Lipinski CA (2004) Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discov Today* 1:337–341
66. van de Waterbeemd H, Gifford E (2003) ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discovery* 2:192–204
67. Wang HY, Cao ZX, Li LL, Jiang PD, Zhao YL, Luo SD, Yang L, Wei YQ, Yang SY (2008) Pharmacophore modeling and virtual screening for designing potential PLK1 inhibitors. *Bioorg Med Chem Lett* 18:4972–4977
68. Schneidman-Duhovny D, Nussinov R, Wolfson HJ (2006) Predicting molecular interactions in silico. II. Protein–protein and protein–drug docking. *Front Med Chem* 3:585–613
69. Taylor RD, Jewsbury PJ, Essex JW (2002) A review of protein–small molecule docking methods. *J Comput Aided Mol Des* 16:151–166
70. Zavodszky MI, Sanschagrin PC, Kuhn LA, Korde RS (2002) Distilling the essential features of a protein surface for improving protein–ligand docking, scoring, and virtual screening. *J Comput Aided Mol Des* 16:883–902
71. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) Extra precision Glide docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J Med Chem* 49:6177–6196
72. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein–ligand docking using GOLD. *Proteins* 52:609–623
73. Tripos Inc. (2004) Sybyl 7.0. Tripos Inc., St. Louis
74. Combes RD, Judson P (2006) The use of artificial intelligence systems for predicting toxicity. *Pestic Sci* 45:179–194
75. Long A, Combes RD (1995) Using DEREK to predict the activity of some carcinogens/mutagens found in foods. *Toxicol In Vitro* 9:563–569
76. Marchant CA, Briggs KA (2008) In silico tools for sharing data and knowledge on toxicity and metabolism: Derek for Windows, Meteor, and Vitic. *Toxicol Mech Methods* 18:177–187
77. Zhang S, Zhang ZY (2007) PTP1B as a drug target: recent developments in PTP1B inhibitor discovery. *Drug Discov Today* 12:373–381
78. Wiesmann C, Barr KJ, Kung J, Zhu J, Erlanson DA, Shen W, Fahr BJ, Zhong M, Taylor L, Randal M (2004) Allosteric inhibition of protein tyrosine phosphatase 1B. *Nat Struct Mol Biol* 11:730–737

Computer-aided de novo ligand design and docking/molecular dynamics study of Vitamin D receptor agonists

Xiu-Long Shen · Midori Takimoto-Kamimura ·
Jing Wei · Qing-Zhi Gao

Received: 26 October 2010 / Accepted: 23 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract $1\alpha,25(\text{OH})_2\text{D}_3$, which is directly mediated by the vitamin D receptor (VDR), exerts a wide variety of biological actions. However, the treatment with $1\alpha,25(\text{OH})_2\text{D}_3$ is limited because of its side effects. Many analogs and several nonsteroidal mimics with potent biological activity have been reported so far, and our rationale for designing the VDR agonists was on the basis of computer-aided drug design method by de novo design of A-ring and C/D-ring position of $1\alpha,25(\text{OH})_2\text{D}_3$. Pyrimidine-2,4-diamine was selected as A-ring, and naphthalene and benzene were chosen as C/D-ring. By linking different components, a virtue compound library was obtained. To evaluate the contribution to activity of each component, we performed a series of automated molecular docking operations. Results revealed that the 19-dimethyl derivatives (the C-19 position correspond to C-20 in $1\alpha,25(\text{OH})_2\text{D}_3$) show the favorable docking affinity to VDR. Moreover, the docking results are quite robust when further

validated by molecular dynamics simulations. In addition, by free energy analysis using molecular mechanics-Poisson-Boltzmann surface area (MM-PBSA) method, the driving force of the binding between VDR and the ligands is proved to be hydrophobic interactions. Thus, a possible strategy to design new series of VDR agonists is proposed. The strategy can be successfully applied to explain the high potential activities of the 19-dimethyl derivatives. It is anticipated that the findings reported here may provide useful information for designing effective VDR agonists as well as the therapeutic treatment of VDR-related diseases.

Keywords De novo design · Molecular docking · Molecular dynamics · VDR agonist · Vitamin D

Introduction

$1\alpha,25$ -Dihydroxyvitamin D_3 ($1\alpha,25(\text{OH})_2\text{D}_3$, 1), the biologically active form of vitamin D_3 , exerts a wide variety of biological actions such as calcium homeostasis, bone mineralization, cellular proliferation and differentiation, immune modulation, and neurobiological functions [1, 2]. Most of these actions are directly mediated by its specific vitamin D receptor (VDR), which is a member of the nuclear receptor superfamily of transcriptional regulators [3, 4]. The binding of the VDR to its specific ligands changes the conformation of the receptor which enables transactivation. This can result in either activation or repression of gene transcription [5, 6].

VDR's transcriptional activity is influenced by several factors such as: binding affinity of ligands, ligand dependent recruitment of co-activators or dissociation of co-repressors [7], efficiency of the ligand uptake into the target cell, tissue specificity, different metabolism of ligands [8],

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1066-8) contains supplementary material, which is available to authorized users.

X.-L. Shen · J. Wei (✉) · Q.-Z. Gao (✉)
Tianjin Key Laboratory for Modern Drug Delivery &
High-Efficiency, School of Pharmaceutical Science
and Technology, Tianjin University,
92 Weijin Road, Nankai District,
Tianjin 300072, People's Republic of China
e-mail: tjucadd@yahoo.com.cn

Q.-Z. Gao
e-mail: qingzhi@gmail.com

M. Takimoto-Kamimura
Teijin Institute for Biomedical Research,
4-3-2 Asahigaoka, Hino,
Tokyo 191-8512, Japan

etc. In this research, the binding affinities of ligands were studied.

$1\alpha,25(\text{OH})_2\text{D}_3$ is used in the treatment of osteoporosis, various types of rickets, secondary hyperparathyroidism, psoriasis, and autoimmune diseases [9]. However, the treatment with **1** is limited because of its hypercalcemia, hypercalciuria and increased bone resorption caused by the required pharmacological doses. As a consequence, more than 3000 analogs and several nonsteroidal mimics of $1\alpha,25(\text{OH})_2\text{D}_3$ with potent biological activity have been synthesized in order to reduce the calcemic side effects [1].

Most of the $1\alpha,25(\text{OH})_2\text{D}_3$ analogs have been reported to show agonistic activity. The analogs are divided into four major classes: the A-ring analogs (e.g., setting frozen A-ring conformation [10]), the seco-B ring analogs (e.g., the 19-nor analogs [11]), the C/D-ring analogs (e.g., the modification of C-9 [12]), and the side chain analogs [1], which most analogs are modified around the side chain. The modifications of the side chain include the addition or transposition of hydroxyl groups [13], the introduction of unsaturation [14], the incorporation of fluorine or deuterium atoms [15], the replacement of a carbon atom with a hetero atom [16], the inversion of the stereochemistry [17], and shortening or lengthening the side chain [18].

The first series of nonsteroidal analogs is characterized by the absence of normal C- and D-rings and by the presence of a five-membered ring, which is named the non C/D-ring series [19]. The bis-phenyl family derivatives have also been reported [9, 20].

In addition, some VDR antagonists have also been studied, such as the lactones [21] and the 26-carboxylic esters [22]. These researches provide a stimulus for the development of new VDR agonists.

Our rationale for designing the VDR agonists was on the basis of computer-aided drug design method by de novo design [23, 24] of A-ring and C/D-ring position of $1\alpha,25(\text{OH})_2\text{D}_3$. Up to now, no reports were found on the use of de novo method for designing and evaluating the potential activity of VDR agonists. A series of novel hits were retrieved by searching the A-ring and C/D-ring fragments of $1\alpha,25(\text{OH})_2\text{D}_3$ and further combining these fragments by the scoring functions and synthetic possibility. Then by molecular docking, each component of the new series of compounds was analyzed and compared systematically to obtain potential VDR agonists with high potential activity. And further MD simulations were performed to investigate the driving forces of the VDR agonists binding to VDR and a possible strategy to design new series of VDR agonists with high potential activity was proposed. It is hoped that the information from this study will provide further understanding of the interaction mechanism and enable

the design of VDR agonists for the therapeutic treatment of VDR-related diseases.

Materials and methods

De novo design

Receptor-based de novo design was performed using InsightII, LUDI module. The LUDI program, developed by Hans-Joachim Böhm [25], positions molecular fragments or small molecules into protein binding sites in such a way that hydrogen bonds are formed with the protein and that lipophilic groups are placed into hydrophobic pockets. The information of protein-ligand binding energies/pockets is very important, not only for revealing true binding mechanisms [26, 27] but also for rational drug design [28, 29]. In order to obtain the desired information for VDR, the three-dimensional structure of VDR ligand binding domain (LBD) was retrieved from the RCSB Protein Data Bank as the de novo design and docking receptor, entry 1DB1.

Compound library design

According to the affinity for VDR, a series of side chains were selected from compound **1** [30], TX522 [31], EB1089 [32], Ro27-5646 [33], EB1213 [17], and MC903 [34]. (All of them are potent VDR agonists whose side chains are named S1-S6, Fig. 2). Considering the synthetic difficulty, the methyl group at the C-21 position was removed from the chiral center C-20 of these side chain analogs.

In order to mimic the structure of **1**, we also suggested four moieties that linked A-ring and C/D-ring fragments by bioisosteric replacement of the seco-B ring of **1**. (named L1-L4, Fig. 2). By combining the different A-ring, linker, C/D-ring and side chain together, a virtual compound library was gained.

Molecular docking

Molecular docking studies were performed using AutoDock, version 4.0. Docking files were prepared using AutoDock Tools v.1.5.2 software. Polar hydrogen atoms were added, non-polar hydrogen atoms were merged and Gasteiger charges and salvation parameters were assigned by default. All docking calculations were made with the Lamarckian genetic algorithm (LGA). A population size of 300 and 25,000,000 energy evaluations were used for 50 search runs. The grid dimensions were $46 \times 46 \times 46$ points with grid spacing of 0.375 \AA . The binding free energies of ligands to the proteins were calculated by using a semiempirical scoring function [35]. The conformations with lowest docked energy were chosen, and analyzed by means of Chimera software.

Molecular dynamics simulations

Based on the docking results, molecular dynamics (MD) simulations for the VDR/VDR agonist (highly potential active compound 36 and natural hormone 1) systems were performed using the SANDER module, implemented in the AMBER 10 package. The standard AMBER ff99SB force field [36] was used as the parameters for the protein, and the general AMBER force field (GAFF) [37] was used as the parameters for the ligands. Sodium counterions were added to neutralize the system. The complexes were then solvated using a shell of TIP3PBOX water model with a closeness parameter of 10 Å away from the boundary of any protein and ligand atoms.

All covalent bonds containing the hydrogen atoms were constrained using SHAKE algorithm [38]. The particle mesh Ewald method [39] was used to treat long-range electrostatic interactions. The cutoff distance for the long-range electrostatic and the van der Waals energy terms were set at 8.0 Å. Energy minimization was used to release the bad contacts in the crystal structure. Subsequently, the complexes of the minimized structures were subjected to 5800 ps MD simulations at 300 K and 1 bar pressure with an integration step of 2 fs and NPT as an ensemble type in our calculations. Trajectories were analyzed using the PTRAJ modules [40].

Free energy analysis by MM-PBSA method

The free energy of ligands binding in the VDR binding pocket was calculated by the molecular mechanics-Poisson-Boltzmann surface area (MM-PBSA) method [41, 42], integrated in AMBER 10. The total free energy of binding is composed of the following terms as Eq. 1:

$$\Delta G_b = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (1)$$

where ΔG_b is the binding free energy in solution; ΔE_{MM} is the molecular mechanics energy, comprised of an electrostatic and van der Waals interaction energies (ΔE_{int}^{ele} and ΔE_{int}^{vdw}) between a ligand and a protein; ΔG_{sol} is the solvation energy, which can be divided into the electrostatic contribution and the hydrophobic contribution to the solvation free energy (ΔG_{sol}^{ele} and ΔG_{sol}^{nonpol}). The electrostatic solvation energy is determined using the finite difference Poisson-Boltzmann (PB) or generalized Born (GB) method, and the nonpolar contribution is estimated by the solvent-accessibility surface method. In this work, the electrostatic contribution to the solvation free energy (ΔG_{sol}^{ele}) was calculated using the GB model [43], while the energy contribution from entropy changes on ligand binding was not included. This was justified by the fact that it is likely that entropy does not contribute much to the relative binding free energies of the ligands to the same protein in several earlier studies applying the same

approximation [44]. The calculated error bars are standard errors of the mean (SE) as Eq. 2:

$$SE = \frac{\text{standard deviation (STD)}}{\sqrt{N}} \quad (2)$$

where N is the number of trajectory snapshots used in the calculations.

Results and discussion

Validation of the docking performance and accuracy

Since there are various docking softwares and scoring functions, it is necessary to validate the docking performance and accuracy of AutoDock [45]. The most straightforward method of evaluating the accuracy of a docking procedure is to determine how closely the lowest energy pose (binding conformation) predicted resembles an experimental binding mode as determined by X-ray crystallography. As cited in literature [46], if the root mean square deviation (RMSD) value of the best-scored conformation is less than 2.0 Å from the experimental one, the prediction is said to be successful. Therefore, the validation of the accuracy of AutoDock 4.0 was done by docking $1\alpha,25(\text{OH})_2\text{D}_3$ into the binding site of VDR. The docking result indicates that the binding conformation of $1\alpha,25(\text{OH})_2\text{D}_3$ superposes excellently with the crystallographical structure as shown in Fig. 1. The RMSD value of the heavy atoms for the top conformation is 0.69 Å. It is clearly noticed that the docked ligand is exactly superimposed on the intact super-agonist 1 with high binding affinity (ΔG_b : -10.32 kcal mol⁻¹). This docking result indicates the high accuracy of the

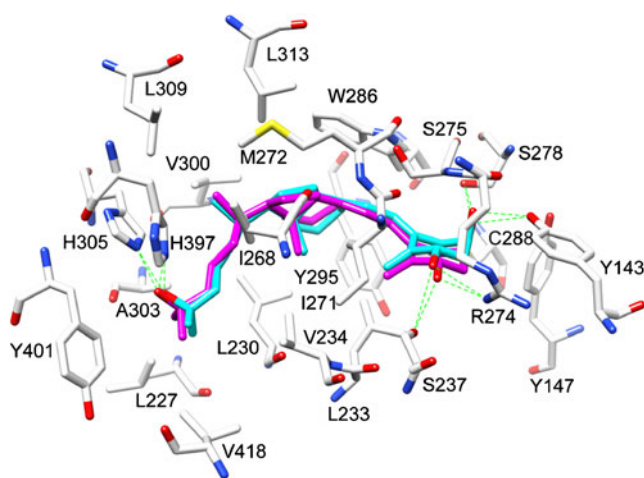


Fig. 1 Docking of the original VDR ligand 1 into its binding site. The intact crystallized VDR ligand is colored in magenta, and the docked ligand is colored in cyan. Their hydrogen bonds with VDR are shown as green dotted lines and the residues in the binding pocket are shown

AutoDock 4.0 in comparison with the biologically bounded crystal structure [47].

The interactions between $1\alpha,25(\text{OH})_2\text{D}_3$ and the receptor VDR involve hydrophobic contacts and electrostatic interactions (Fig. 1). The A ring of the natural hormone has hydrophobic interactions with C288 and Y147. The seco-B to C ring part is sandwiched by the side chains of hydrophobic residues, L233, V234, I268, I271, M272, S275, W286, Y295 and L313. The hydrophobic D ring is positioned in a hydrophobic pocket formed by the side chains of L227, L230, V300, A303, L309, Y401 and V418. In addition, the docked ligand exhibits six true hydrogen bonds with S237 (OG, 3.277 Å), R274 (N₁H, 3.157 Å), Y143 (OH, 2.754 Å), S278 (OG, 2.892 Å), H305 (NE2, 2.871 Å), and H397 (NE2, 3.051 Å), which are exactly the same amino acids as those found in VDR and ligand co-crystal structure. Among these residues related to hydrogen bonds, Y143, R274 and H397 are reported to play an important role to anchor the ligand into the pocket to show the agonistic activity [48].

De novo design and compound library design

The center of search parameter was defined by the A-ring and C/D-ring of the natural ligand, 1. The default fragment library (available with InsightII software) and Energy_Estimate_1 scoring function were used to design and estimate the fragments. All other parameters were assigned by default. By de novo design, we obtained 51 A-ring fragments and 33 C/D-ring fragments. The structures and scores of A-ring and C/D-ring fragments are shown in Fig. S1 and Table S1, Fig. S2 and Table S2, respectively. (Supplementary material) The possible connecting positions of the fragments were also defined according to their orientations in the pocket.

In terms of A-ring fragment, pyrimidine-2,4-diamine (A1) was chosen due to its high score (total score ranks 4th and the H-Bond score ranks 1st among the 51 fragments), synthetic possibility and similarity to A-ring of 1. As to C/D-ring fragments, benzene (CD1) and naphthalene (CD2) were chosen for our study because of their synthetic possibility and similarity to C/D-ring of 1, though their scores are not among the top ones. By combining the different A-ring, Linker, C/D-ring and Side chain fragments described before, we gained a primary $1 \times 4 \times 2 \times 6$ virtual compound library. The fragments of each components and the structure mode of the new series of VDR agonists are shown in Fig. 2.

In silico screening by evaluation of each component

The virtual compound library was then screened by AutoDock and the structure modes and docking results of

some of these compounds used in this section are shown in Table 1.

Evaluation of the C/D-ring fragments

Compounds containing naphthalene (CD2) at the C/D-ring position reveal the higher binding affinities than benzene (CD1). The lowest binding energies of the former are obtained within the range of -8.72 to -8.85 kcal mol⁻¹ (compounds 6–9), while this range of the latter is -6.81 to -7.22 kcal mol⁻¹ (compounds 2–5). This favored docking affinity of naphthalene at the C/D-ring position is mainly attributed to the stronger hydrophobic interactions with the hydrophobic residues in the pocket, such as L230, V234, I271, M272, W286 and Y295.

Evaluation of the side chains

In general, compounds with S3-S5 (12–17) as side chains exhibit lower binding docking energies (<-9.5 kcal mol⁻¹) than compounds with S1, S2 and S6 (6–7, 10–19). The differences of the van der Waals, H-bond and desolvation energy (E_{vdw}) of these compounds may explain this phenomenon. The E_{vdw} of compounds with S3-5 is within the range of -12.39 to -13.17 kcal mol⁻¹, whereas this range of compounds with S1, S2 and S6 is -10.53 to -11.31 kcal mol⁻¹. The ethyl groups, the double-branched chain and the benzene ring of compounds with S3-5 may account for the lower E_{vdw} because of their enhanced hydrophobic contacts with the receptor. In addition, the estimated binding energies for compounds 12, 16, 17 obtained in this study are lower than that of the natural hormone.

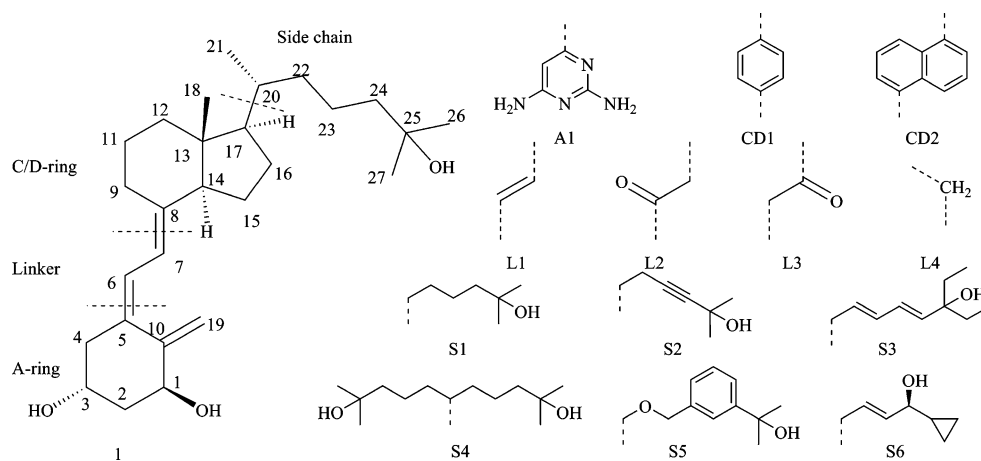
Evaluation of the linkers

In general, there are no obvious differences in terms of binding energies among the compounds with L1-L4 as linker between A-ring and C/D-ring. According to the binding energy, the compounds with L1 exhibit slightly enhanced docking affinities compared to the compounds with L2-4, see Table 1 (6–9, 12–17, 20–25). This may be explained by the proton- π interaction between the conjugated triene part of compounds with A1+L1+CD2 fragments and the residues W286 and S275 in the pocket, parallel to the proton- π interaction between the triene conjugated part of 1 with these two residues [48, 49].

Structure optimization

From the evaluation above, compounds with CD2, S3, S5 and L1 were selected for the study of structure optimization because they exhibit higher affinity with the receptor (Table 2).

Fig. 2 The structure modes and the components of the VDR agonists involved in this study



In fact, there are four possible linking positions of CD2-naphthalene with the linker and side chain according to their orientations in the pocket since this pocket is quite large for our ligands. As shown in Fig. 3, in spite of the 1,5-disubstituted naphthalene compounds (compounds 12, 16), there are also 2,8-disubstituted naphthalene compounds (26, 29), 2,7-disubstituted naphthalene compounds (27, 30) and 1,6-disubstituted naphthalene compounds (28, 31). The

compounds with 1,5-, 2,7-, and 1,6-disubstituted naphthalene are favorable for the binding affinity into VDR as shown in the binding free energies, e.g. 16 ($-10.75 \text{ kcal mol}^{-1}$), 27 ($-10.5 \text{ kcal mol}^{-1}$), and 31 ($-10.61 \text{ kcal mol}^{-1}$), whereas the lower binding affinities of derivatives with 2,8-disubstituted naphthalene such as 26 ($-10.03 \text{ kcal mol}^{-1}$) and 29 ($-9.94 \text{ kcal mol}^{-1}$) are revealed. This may mainly be attributed to the flexibility of the side chains, which contain

Table 1 The structure modes and docking results of compounds 1-25

Entry	A-ring	Link	C/D-ring	Side chain	E_{binding} (kcal/mol)	E_{vdw} (kcal/mol)	E_{ele} (kcal/mol)
1	-	-	-	-	-10.32	-12.27	-0.32
2	A1	L1	CD 1	S 1	-7.22	-9.3	-0.15
3	A1	L2	CD 1	S 1	-7.14	-9.71	-0.11
4	A1	L3	CD 1	S 1	-7.15	-9.66	-0.19
5	A1	L4	CD 1	S 1	-6.81	-9.05	-0.16
6	A1	L1	CD 2	S 1	-8.85	-11.01	-0.26
7	A1	L2	CD 2	S 1	-8.74	-11.31	-0.09
8	A1	L3	CD 2	S 1	-8.84	-11.16	-0.18
9	A1	L4	CD 2	S 1	-8.72	-11.2	-0.21
10	A1	L1	CD 2	S 2	-8.95	-10.64	-0.23
11	A1	L2	CD 2	S 2	-8.29	-10.53	-0.24
12	A1	L1	CD 2	S 3	-10.37	-12.63	-0.21
13	A1	L2	CD 2	S 3	-9.68	-12.39	-0.11
14	A1	L1	CD 2	S 4	-9.88	-13.11	-0.02
15	A1	L2	CD 2	S 4	-10.05	-13.39	-0.33
16	A1	L1	CD 2	S 5	-10.75	-12.72	-0.27
17	A1	L2	CD 2	S 5	-10.41	-12.71	-0.13
18	A1	L1	CD 2	S 6	-8.83	-10.69	-0.14
19	A1	L2	CD 2	S 6	-8.45	-10.93	0.04
20	A1	L3	CD 2	S 3	-10.22	-12.64	-0.3
21	A1	L4	CD 2	S 3	-8.96	-11.5	-0.05
22	A1	L3	CD 2	S 4	-9.35	-12.83	-0.12
23	A1	L4	CD 2	S 4	-9.73	-12.62	-0.03
24	A1	L3	CD 2	S 5	-10.43	-12.82	-0.09
25	A1	L4	CD 2	S 5	-10.17	-12.32	-0.25

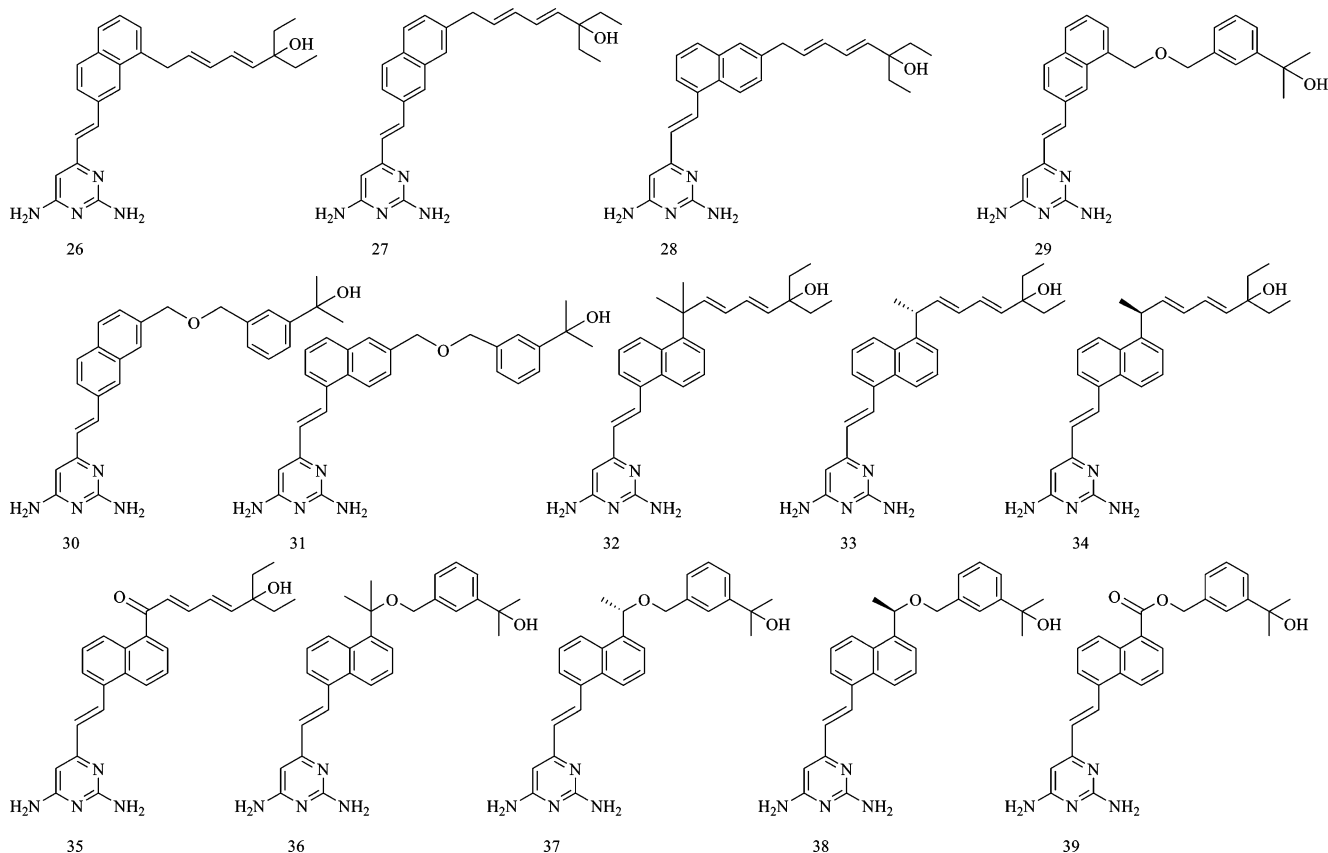
Table 2 The docking results of compounds 26–39 in the structure optimization study

Entry	E_{binding} (kcal/mol)	E_{vdw} (kcal/mol)	E_{ele} (kcal/mol)
26	-10.03	-12.33	-0.21
27	-10.5	-12.71	-0.22
28	-10.39	-12.63	-0.23
29	-9.94	-11.87	-0.22
30	-10.43	-12.36	-0.23
31	-10.61	-12.54	-0.21
32	-10.98	-13.04	-0.21
33	-10.64	-12.89	-0.21
34	-10.34	-12.46	-0.17
35	-10.7	-13.05	-0.21
36	-11.5	-13.73	-0.12
37	-10.82	-12.76	-0.25
38	-10.06	-12.46	-0.05
39	-10.85	-12.93	-0.18

single bonds that can rotate in order to attain favorable interactions with VDR, and thus leads to several good linking positions of naphthalene. However, taking synthetic possibility into consideration, the 1,5-disubstituted naphtha-

lene linking position may be the best choice to obtain compounds with satisfying potential activity.

As cited in literature [9], structure-activity studies of vitamin D analogs show that the stereochemistry at C-20 of $1\alpha,25(\text{OH})_2\text{D}_3$ is the key structural motifs for vitamin D actions. Therefore, by changing the stereochemistry of the corresponding position, C-19 in our analogs, we investigated its effects on binding energy. As shown in Fig. 3, we designed 19-dimethyl derivatives (32, 36), 19-natural derivatives (33, 37), 19-epi derivatives (34, 38), and 19-oxo derivatives (35, 39). The 19-dimethyl, 19-natural and 19-oxo derivatives exhibit slightly higher binding affinity than the 19-epi derivatives and 20-nor derivatives (12, 16). The lower binding energies of 19-dimethyl analogs are mainly attributed to the enhanced hydrophobic interactions between the ligands and V300, L309 and L313 of the binding site. And the high binding affinity of 19-natural analogs can be explained by the stronger hydrophobic interaction between these three residues and the methyl group at C-19 position 20-nor analogs which is important in ligand harboring and helps overcome the unfavorable conformation at C-(16-17-20-22) of the side chain of $1,25(\text{OH})_2\text{D}_3$. The low affinity of 19-epi analogs is due to substantial reorientation of the side chain by epimeriza-

**Fig. 3** The structures of compound 26–39 in the structure optimization study

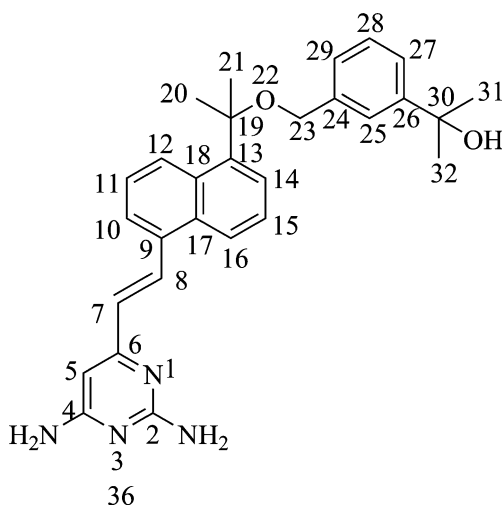


Fig. 4 The numbered structure of compound 36

tion, which interferes with their surface interaction into the pocket of VDR.

Molecular dynamics (MD) studies of the 19-dimethyl derivatives

In order to verify if the results obtained here by docking are robust or a fortuitous results, we carried out two 5.8 ns molecular dynamics simulations with one highly potential active compound of the 19-dimethyl derivatives (36, the numbered structure was shown in Fig. 4) by molecular docking study and the natural hormone, 1.

To gauge whether the 5.8 ns MD simulations were stable and converged, the atomic RMSDs of the C_{α} , C, and N atoms of the protein and the ligands obtained during the trajectories and the initial structures were monitored during the simulations (Fig. 5). The series of snapshots between

0.8 and 5.8 ns of the equilibrium phase was used for the calculations of RMSD. It can be seen that the two complexes become stable after about 1 ns in the last 5 ns trajectories with a mean RMSD value of 1.36 Å for the backbone of the receptor, 0.50 Å for the ligand in the VDR/compound 1 system and 1.53 Å for the backbone of the receptor, 0.62 Å for the ligand in the VDR/compound 36 system. Both the two protein-ligand trajectories exhibit low backbone and ligand RMSD values, which indicate the high binding affinities of VDR with the natural hormone 1 and compound 36, further enhancing the credibility of the docking results.

Further insight into the forces involved in substrate binding can be obtained by analysis of the MM/GBSA free energies contributions, which are listed in Table 3 for the two ligands. The series of snapshots between 3.8 and 5.8 ns of the equilibrium phase was used for free energy calculations, taken at 2 ps intervals from the trajectories of each simulation. The calculated $\Delta G_{binding}$ for the VDR/compound 36 ($-59.01 \text{ kcal mol}^{-1}$) is slightly higher than the value of the VDR/natural hormone 1 ($-58.43 \text{ kcal mol}^{-1}$), demonstrating that the potential activity of compound 36 may be as high as the natural hormone.

Both the intermolecular van der Waals and the electrostatics interactions are significant contributions to the binding, whereas polar solvation terms counteract binding. Nonpolar solvation terms, which correspond to the burial of solvent-accessible surface-area upon binding, contribute slightly favorably. Though the gas-phase electrostatic values, ΔE_{int}^{ele} of the two complexes, show that electrostatic interactions are in favor of the binding, the overall electrostatic interactions energies, ΔG_{ele} ($\Delta E_{int}^{ele} + \Delta E_{solv}^{ele}$), are positive and unfavorable for the binding, which may be caused by the large desolvation penalty of charged and

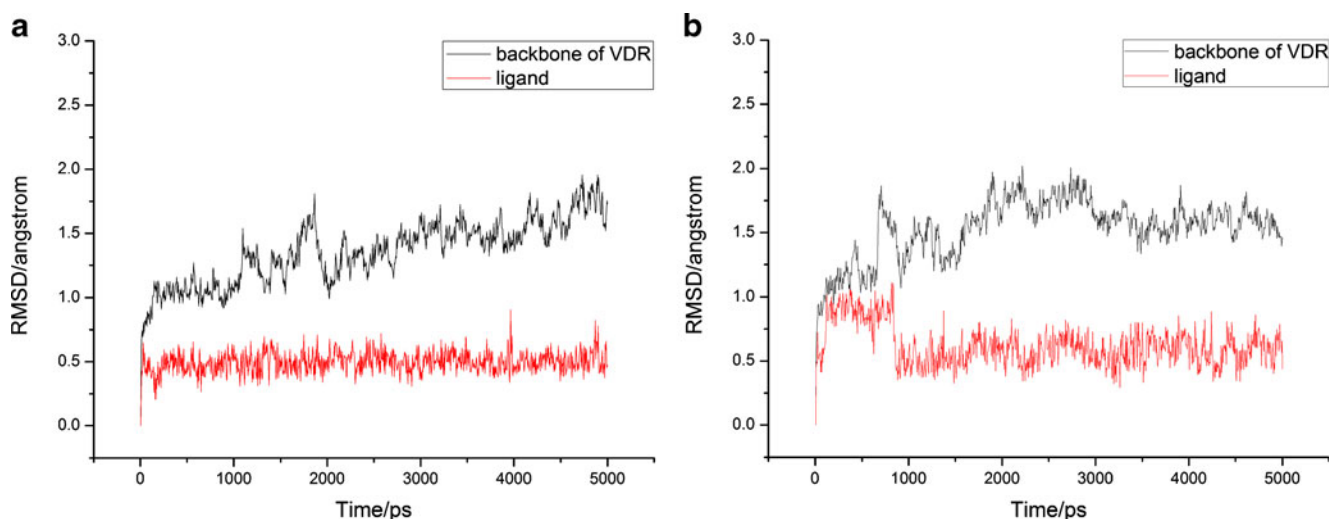


Fig. 5 Root mean square deviation (RMSD) obtained during the last 5000 ps MD simulations for the backbone of VDR and (a) natural hormone 1, (b) compound 36. The deviation is with respect to the corresponding starting structure as a function of time

Table 3 Individual energy components for the calculated binding free energies of VDR and the two VDR agonists, natural hormone 1 and compound 36 using MM_GBSA method

Contribution	VDR/compound 1	VDR/compound 36
$\Delta E_{\text{int}}^{\text{ele}}$	-23.77 (0.31)	-17.47 (0.24)
$\Delta E_{\text{int}}^{\text{vdw}}$	-62.17 (0.25)	-70.03 (0.23)
$\Delta G_{\text{sol}}^{\text{nonpol}}$	-8.30 (0.01)	-8.55 (0.01)
$\Delta G_{\text{sol}}^{\text{ele}}$	35.80 (0.19)	37.03 (0.15)
$\Delta G_{\text{sol}}^{\text{a}}$	27.51 (0.19)	28.49 (0.15)
$\Delta G_{\text{ele}}^{\text{b}}$	12.03 (0.27)	19.56 (0.19)
ΔG_{b}	-58.43 (0.30)	-59.01 (0.22)

^a The polar/nonpolar ($\Delta G_{\text{sol}}^{\text{nonpol}}$ + $\Delta G_{\text{sol}}^{\text{ele}}$) contributions

^b The electrostatic ($\Delta E_{\text{int}}^{\text{ele}}$ + $\Delta G_{\text{sol}}^{\text{ele}}$) contributions. All energies are averaged over 200 snapshots and are given in kcal mol⁻¹. Calculation of ΔG_{b} does not explicitly consider entropy contributions. The values in parentheses represent the standard error of the mean

polar groups that is not sufficiently compensated by complex formation. Comparing the van der Waals/nonpolar ($\Delta E_{\text{int}}^{\text{vdw}}$ + $\Delta G_{\text{sol}}^{\text{nonpol}}$) contributions with the electrostatic contributions ΔG_{ele} , we notice that the association between the VDR and compounds 1 and 36 is mainly driven by more favorable van der Waals/nonpolar interaction in the complex than in solution. The electrostatic interactions between VDR and compounds 1 and 36 are strong, but the electrostatic interactions between the solvent (water molecules) and the ligand are much stronger, thus the total electrostatic interactions inhibit the ligand binding.

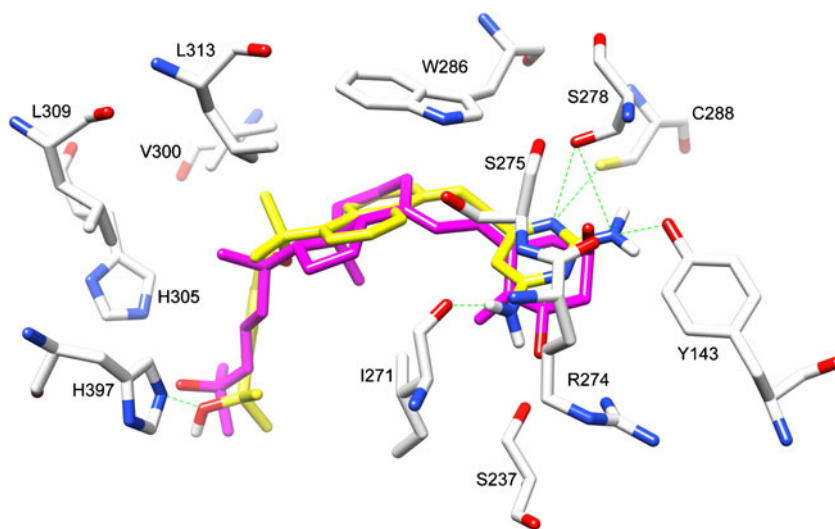
Therefore, the fact that the van der Waals/nonpolar contribution of the VDR/compound 36 is 8.11 kcal mol⁻¹ lower than that of the VDR/compound 1 systems reveals that compound 36 fits more snugly within the active-site cavity and has a slight tighter binding to VDR than the natural hormone, adding nonpolar packing in the active site

to a deeper extent. And in terms of the $\Delta E_{\text{int}}^{\text{ele}}$ contributions, VDR/natural hormone 1 complex exhibits 6.3 kcal mol⁻¹ lower value than that of VDR/compound 36 complex. This is because compound 36 loses some important hydrogen bonds. Shown in Fig. 6, compound 36 enjoys six hydrogen bonds with Y143, I271, S278 (two hydrogen bonds), C288 and H397, which loses hydrogen bonds with S237, R274 and H305 of the natural hormone 1 and has another three hydrogen bonds with I271, S278 and C288 because of the heterocyclic pyrimidine-2,4-diamine as A-ring. Taking all energy contributions into account, compound 36 presents slightly higher potential activity than 1. Hence, inspired by the molecular dynamics studies, we conclude a possible strategy to design new series of VDR agonists with high potential activity, which is to retain the important hydrogen bonds and enhance the hydrophobic interactions with the hydrophobic residues in the pocket by adding the hydrophobic moieties to modify the original structures. In our study, 19-dimethyl derivatives exhibit the most favorable affinity with VDR. The reason is that 19-dimethyl derivatives obey the strategy above that they retain two important hydrogen bonds with Y143 and H397 and increase the hydrophobic interactions between the side chains and the receptor.

Conclusions

Through the de novo design of the A-ring and C/D-ring of the natural ligand, 1 α ,25(OH)₂D₃, some fragments were gained and opted. And combining the selected one A-ring, four Linkers, two C/D-rings and six side chains fragments, a primary 1 \times 4 \times 2 \times 6 virtual compound library was built. By in silico screening, each component was evaluated according to binding free energy from AutoDock and the

Fig. 6 Comparison of the docked compound 36 with position of 1 in the reference complex 1DB1. The intact crystallized VDR ligand is colored in magenta, and the docked compound 36 is colored in yellow. The hydrogen bonds of 36 with VDR are shown as green dotted lines



compounds with CD2, S3, S5 and L1 were chosen for the study of structure optimization. Compounds with different linking positions of naphthalene as C/D-ring were tested and the 1,5-disubstituted naphthalene linking position may be the best way to obtain compounds with high potential activity. In addition, we also investigated the necessity of the stereochemistry at the C-19 position of our analogs and the results show that the methyl group at the C-19 position attributes mainly to the hydrophobic contacts with the receptor. Therefore, we suggested the 19-dimethyl derivatives which consequently show the most favorable docking affinity to VDR among the designed compounds.

In order to achieve further understanding of the driving forces in the association of VDR and VDR agonists, MD simulations of VDR and two ligands, compound 1 and 36 were performed and the free energy was calculated by MM-GBSA method. The free energy analysis demonstrates that the binding between VDR and compound 1 and 36 is mainly driven by more favorable van der Waals/nonpolar interaction in the complex than in solution. And by comparison of the free energies of compound 1 and 36, the high potential activity of the 19-dimethyl derivatives is explained and a possible strategy to design a new series of VDR agonists is proposed. The strategy is to retain the important hydrogen bonds and increase the hydrophobic interactions with the hydrophobic residues in the binding site. The 19-dimethyl derivatives may be worth further study and it is anticipated that the findings reported here may provide useful information for designing effective VDR agonists for the therapeutic treatment of VDR-related diseases.

Acknowledgments We gratefully acknowledge the support from Project of Undergraduate Educational Reform & Capacities in Tianjin University (No. 200904050).

References

- Bouillon R, Okamura WH, Norman AW (1995) Structure-function relationships in the vitamin D endocrine system. *Endocr Rev* 16:200–257
- Norman AW, Frankel JB, Heldt AM, Grodsky GM (1980) Vitamin D deficiency inhibits pancreatic secretion of insulin. *Science* 209:823–825
- Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schutz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P, Evans RM (1995) The nuclear receptor superfamily: the second decade. *Cell* 83:835–839
- Nakano Y, Kato Y, Imai K, Ochiai E, Namekawa J, Ishizuka S, Takenouchi K, Tanatani A, Hashimoto Y, Nagasawa K (2006) Practical synthesis and evaluation of the biological activities of 1 α ,25-dihydroxyvitamin D₃ antagonists, 1 α ,25-dihydroxyvitamin D₃-26,23-lactams. Designed on the basis of the helix 12-folding inhibition hypothesis. *J Med Chem* 49:2398–2406
- Evans RM (1988) The steroid and thyroid hormone receptor superfamily. *Science* 240:889–895
- Peleg S, Posner GH (2003) Vitamin D analogs as modulators of vitamin D receptor action. *Curr Top Med Chem* 3:1555–1572
- Aranda A, Pascual A (2001) Nuclear hormone receptors and gene expression. *Physiol Rev* 81:1269–1304
- Brown AJ (2000) Mechanisms for the selective actions of vitamin D analogues. *Curr Pharm Des* 6:701–706
- Masuda S, Jones G (2003) Vitamin D analogs-drug design based on proteins involved in vitamin D signal transduction. *Curr Drug Targets Immune Endocr Metab Disord* 3:43–66
- Sicinski RR, Glebocka A, Plum LA, DeLuca HF (2007) Design, synthesis, and biological evaluation of a 1 α ,25-Dihydroxy-19-norvitamin D₃ analogue with a frozen a-ring conformation. *J Med Chem* 50:6154–6164
- Bouillon R, Sarandeses LA, Allewaert K, Zhao J, Mascarenas JL, Mourino A, Vrielynck S, de Clercq P, Vandewalle M (1993) Biologic activity of dihydroxylated 19-nor-(pre)vitamin D₃. *J Bone Miner Res* 8:1009–1015
- Okamura WH, Aurrecoechea JM, Gibbs RA, Norman AW (1989) Synthesis and biological activity of 9,11-Dehydrovitamin D₃ analogues: stereoselective preparation of 6 β -Vitamin D vinylallenes and a concise enynol synthesis for preparing the A-ring. *J Org Chem* 54:4072–4083
- Wang YZ, Li H, Bruns ME, Uskokovic M, Truitt GA, Horst R, Reinhardt T, Christakos S (1993) Effect of 1,25,28-trihydroxyvitamin D₂ and 1,24,25-trihydroxyvitamin D₃ on intestinal calbindin-D_{9k} mRNA and protein: Is there a correlation with intestinal calcium transport? *J Bone Miner Res* 8:1483–1490
- Riveiros R, Rumbo A, Sarandeses LA, Mourino A (2007) Synthesis and Conformational Analysis of 17 α ,21-Cyclo-22-Unsaturated Analogues of Calcitriol. *J Org Chem* 72:5477–5485
- Yamada S, Yamamoto K, Masuno H, Ohta M (1998) Conformation-function relationship of vitamin D: conformational analysis predicts potential side-chain structure. *J Med Chem* 41:1467–1475
- Westermann J, Schneider M, Platzeck J, Petrov O (2007) Practical synthesis of a heterocyclic immunosuppressive vitamin D analogue. *Org Process Res Dev* 11:200–205
- Bury Y, Ruf D, Hansen CM, Kissmeyer AM, Binderup L, Carlberg C (2001) Molecular evaluation of vitamin D₃ receptor agonists designed for topical treatment of skin diseases. *J Invest Dermatol* 116:785–792
- Inaba Y, Yoshimoto N, Sakamaki Y, Nakabayashi M, Ikura T, Tamamura H, Ito N, Shimizu M, Yamamoto K (2009) A new class of vitamin D analogues that induce structural rearrangement of the ligand-binding pocket of the receptor. *J Med Chem* 52:1438–1449
- Verstuyf A, Verlinden L, van Etten E, Shi L, Wu Y, D'Halleweyn C, van Haver D, Zhu GD, Chen YJ, Zhou X, Haussler MR, De Clercq P, Vandewalle M, van Baelen H, Mathieu C, Bouillon R (2000) Biological activity of CD-ring modified 1 α ,25-dihydroxyvitamin D analogues: C-ring and five-membered D-ring analogues. *J Bone Miner Res* 15:237–252
- Hosoda S, Tanatani A, Wakabayashi K, Makishima M, Imai K, Miyachi H, Nagasawa K, Hashimoto Y (2006) Ligands with a 3,3-diphenylpentane skeleton for nuclear vitamin D and androgen receptors: dual activities and metabolic activation. *Bioorg Med Chem* 14:5489–5502
- Miuraa D, Manabeb K, Gao QZ, Normanc AW, Ishizuka S (1999) 1 α ,25-Dihydroxyvitamin D₃-26,23-lactone analogs antagonize differentiation of human leukemia cells (HL-60 cells) but not of human acute promyelocytic leukemia cells (NB4 cells). *FEBS Lett* 460:297–302
- Bury Y, Steinmeyer A, Carlberg C (2000) Structure activity relationship of carboxylic ester antagonists of the vitamin D(3) receptor. *Mol Pharmacol* 58:1067–1074
- García-Sosa AT, Mancera RL (2006) The effect of a tightly bound water molecule on scaffold diversity in the computer-aided de novo ligand design of CDK2 inhibitors. *J Mol Model* 12:422–431

24. Francis SM, Mittal A, Sharma M, Bharatam PV (2008) Design of Benzene-1,2-diamines as selective inducible nitric oxide synthase inhibitors: a combined de novo design and docking analysis. *J Mol Model* 14:215–224
25. Böhm HJ (1994) On the use of LUDI to search the fine chemicals directory for ligands of proteins of known three-dimensional structure. *J Comput Aid Mol Des* 8:623–632
26. Huang RB, Du QS, Wang CH, Chou KC (2008) An in-depth analysis of the biological functional studies based on the NMR M2 channel structure of influenza A virus. *Biochem Biophys Res Commun* 377:1243–1247
27. Du QS, Huang RB, Wang CH, Li XM, Chou KC (2009) Energetic analysis of the two controversial drug binding sites of the M2 proton channel in influenza A virus. *J Theor Biol* 259:159–164
28. Wang SQ, Du QS, Chou KC (2007) Study of drug resistance of chicken influenza a virus (H5N1) from homology-modeled 3D structures of neuraminidases. *Biochem Biophys Res Commun* 354:634–640
29. Wang SQ, Du QS, Huang RB, Zhang DW, Chou KC (2009) Insights from investigating the interaction of oseltamivir (Tamiflu) with neuraminidase of the 2009 H1N1 swine flu virus. *Biochem Biophys Res Commun* 386:432–436
30. Norman AW, Myrtle JF, Miogett RJ, Nowicki HG, Williams V, Popjaák G (1971) 1,25-Dihydroxycholecalciferol: identification of the proposed active form of vitamin D3 in the intestine. *Science* 173:51–54
31. Verlinden L, Verstuyf A, Van Camp M, Marcelis S, Sabbe K, Zhao XY, De Clercq P, Vandewalle M, Bouillon R (2000) Two novel 14-epi-analogues of 1,25-dihydroxyvitamin D₃ inhibit the growth of human breast cancer cells in vitro and in vivo. *Cancer Res* 60:2673–2679
32. Hansen CM, Hamberg KJ, Binderup E, Binderup L (2000) Seocalcitol (EB 1089) A vitamin D analogue of anti-cancer potential. Background, design, synthesis, pre-clinical and clinical evaluation. *Curr Pharm Des* 26:803–828
33. Hisatake J, O'Kelly J, Uskokovic MR, Tomoyasu S, Koeffler HP (2001) Novel vitamin D₃ analog, 21-(3-methyl-3-hydroxy-butyl)-19-nor D₃, that modulates cell growth, differentiation, apoptosis, cell cycle, and induction of PTEN in leukemic cells. *Blood* 97:2427–2433
34. Kissmeyer AM, Binderup L (1991) Calcipotriol (MC 903): pharmacokinetics in rats and biological activities of metabolites. A comparative study with 1,25(OH)₂D₃. *Biochem Pharmacol* 41:1601–1606
35. Huey R, Morris GM, Olson AJ, Goodsell DS (2007) A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* 28:1145–1152
36. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197
37. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* 25:1157–1174
38. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23:327–341
39. Darden T, York D, Pedersen L (1993) Particle mesh ewald: an N-log(N) method for ewald sums in large systems. *J Chem Phys* 98:10089–10092
40. Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. *Science* 268:1144–1149
41. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33:889–897
42. Massova I, Kollman PA (2000) Combined molecular mechanical and continuum solvent approach (MM PBSA/GBSA) to predict ligand binding. *Perspect Drug Discov Des* 18:113–135
43. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55:383–394
44. Zhou ZG, Madrid M, Evansck JD, Madura JD (2005) Effect of a bound non-nucleoside RT inhibitor on the dynamics of wild-type and mutant HIV-1 reverse transcriptase. *J Am Chem Soc* 127:17253–17260
45. Bissantz C, Folkers G, Rognan D (2000) Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* 43:4759–4767
46. Wang R, Lu Y, Wang S (2003) Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem* 46:2287–2303
47. Plummer MS, Holland DR, Shahripour A, Lunney EA, Fergus JH, Marks JS, McConnell P, Mueller WT, Sawyer TK (1997) Design, synthesis, and cocrystal structure of a nonpeptide Src SH2 domain ligand. *J Med Chem* 40:3719–3725
48. Yamamoto K, Abe D, Yoshimoto N, Choi M, Yamagishi K, Tokiwa H, Shimizu M, Makishima M, Yamada S (2006) Vitamin D receptor: ligand recognition and allosteric network. *J Med Chem* 49:1313–1324
49. Yamagishia K, Tokiwaa H, Makishimac M, Yamada S (2010) Interactions between 1 α ,25(OH)₂D₃ and residues in the ligand-binding pocket of the vitamin D receptor: a correlated fragment molecular orbital study. *J Steroid Biochem* 121:63–67

Simulated Q-annealing: conformational search with an effective potential

Won-joon Son · Soonmin Jang · Seokmin Shin

Received: 19 November 2010 / Accepted: 24 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract We have tested a version of the generalized simulated annealing algorithm based on molecular dynamics simulations with effective potential suggested by Tsallis statistics. The generalized annealing method, termed “simulated Q-annealing (SQ),” is applied to the simulations of a synthetic 11-residue peptide segment (1AQG). In SQ, the energy barriers between local minima change as the parameter q is varied and specific degrees of freedom can be selectively heated up and annealed. Conformational dynamics obtained by ordinary simulated annealing (SA) and SQ simulations are compared in order to illustrate the effectiveness of the SQ approach in conformational searching. We show that SQ can navigate the potential energy surface efficiently with a simple annealing protocol and demonstrate that conformations sampled by SQ can represent the funnel-like free energy surface.

Keywords Conformation search · Effective potential · Molecular dynamics simulations · Protein folding · Simulated annealing

Introduction

Among the early heuristics for the multiple minima problems in optimization for molecular simulation studies, simulated annealing (SA) [1, 2] and its derivatives have been one of the

customary methods. In particular, the robustness of SA in searching for potential energy surfaces to determine the ground state configuration has been exploited since its early stage [3]. The main idea of SA is to reproduce a process of making a single crystal from a melt by careful annealing. However, as is already noted in the original paper [2], a rapid cooling or simple minimization scheme may result in only metastable states far from the accessible optimal solution. Since SA is a heuristic method in its strict sense, details of the annealing protocol (including temperature change scheme, number of steps, termination criteria, *etc.*) affect the final solution significantly [4]. In particular, it is not always clear how to judge the convergence of SA, even if the adopted cooling schedule ensures the best solution [5]. Thus, a reliable optimal solution is accessible only when the series of tailored simulations are performed with close monitoring. It was also reported by Geman, et al. [6] that the annealing temperature in logarithmic (with respect to simulation time) fashion is required in the original SA with Boltzmann-Gibbs distribution in order to find the global minimum of the given optimization problem. However, this argument will not be strictly applied to generalized simulated annealing algorithms using non-Boltzmann statistics. Among those non-Boltzmann statistics suggested, a generalized simulated annealing algorithm based on Tsallis statistics was shown to be more efficient than other classical simulated annealing and fast simulated annealing algorithms, especially as the complexity of the system increased [7]. In the early generalized simulated annealing algorithms, the Tsallis-Stariolo form of the Cauchy-Lorentz visiting distribution is used in place of the Gaussian visiting distribution [8]. Andricioaei and Straub have suggested another generalized acceptance probability which obeys the detailed balance condition [9].

Several methods for improving annealing algorithms, including stochastic tunneling, energy landscape paving

W.-j. Son · S. Shin (✉)
School of Chemistry, Seoul National University,
Seoul 151-747, Korea
e-mail: sshin@snu.ac.kr

S. Jang
Department of Chemistry, Sejong University,
Seoul 143-747, Korea

and basin hopping methods, have been attempted in simulating conformational dynamics of proteins [10–15]. In this paper, a molecular-dynamics version of the generalized simulated annealing algorithm based on Tsallis statistics, termed “simulated Q-annealing (SQ)”, is applied to the simulations of model peptides. Conformational dynamics obtained using the SA and SQ simulations are compared in order to illustrate the efficiency of the SQ approach in conformational searching.

Methods

The simulated annealing (SA) approach for complex molecules like proteins is commonly performed with a molecular dynamics (MD) algorithm because the generation of new configurations in naïve Monte Carlo schemes may not be efficient for molecules of complex geometry. One way of incorporating SA and MD is to integrate the equations of motion while the system temperature is controlled through a coupled heat bath [16]. The generalized simulated annealing proposed by Andricioaei and Straub [9] was coupled with a generalized molecular dynamics with the effective potential obtained from Tsallis transformation [17]. To sample the trajectories compatible with the Tsallis probability distribution, the equations of motion based on the effective potential were exploited. For a particle of mass m_k and position \vec{r}_k moving under potential U , the equation of motion becomes

$$m_k \frac{d^2 \vec{r}_k}{dt^2} = -\nabla_{\vec{r}_k} \bar{U} = -\left(\frac{q}{1 - (1 - q)\beta U(\vec{r}^N)} \right) \nabla_{\vec{r}_k} U(\vec{r}^N), \quad (1)$$

where

$$\bar{U} = \frac{q}{\beta(q-1)} \ln[1 - (1 - q)\beta U]. \quad (2)$$

β is defined to be $1/k_B T$, where k_B is the Boltzmann constant and T is the temperature of the system, respectively. When U is negative, a positive constant ε should be added to U , so that the argument of the logarithm function in Eq. 2 should not lead to divergence. Note that if ε is too large, the overall behavior of the scaling factor may be dominated by ε . Even when $U(\vec{r}^N)$ is negative, the denominator of the scaling factor in Eq. 1 can be defined and the equation of motion can generate dynamics trajectories. However, $\nabla_{\vec{r}_k} \bar{U}$ is not defined in that case and the resulting trajectories are of no meaning. As can be

seen, the particle feels the potential energy scaled by a factor of $q[1 - (1 - q)\beta U(\vec{r}^N)]^{-1}$, effectively. Put in other words, q increases or decreases the effective temperature of the system, $1/\beta' = k_B T'$, such that [17]

$$\frac{1}{\beta'} = \frac{1}{\beta} - (1 - q)\varepsilon. \quad (3)$$

In the original version of SA, an effective temperature defines the annealing space. In the molecular dynamics version of the generalized simulated annealing on the q -space, denoted as “simulated Q-annealing” (SQ) hereafter, q takes the place of the temperature in the original SA. The parameter q may be considered to be the effective temperature. In SQ, this effective temperature lowers the potential energy barriers rather than increasing the kinetic energy of each particle. The parameter q is changed from a certain value bigger than one, which enables conformation samplings in highly smoothed potential energy surfaces, to $q=1$, which corresponds to the original potential energy surface. In SQ, the system temperature is fixed throughout the annealing by the usual thermostat method.

Because the energy barriers between local minimum conformations are lowered when q is larger than 1, the trajectory trapping problem can be reduced. Also, specific degrees of freedom in the force component can be selectively *heated up* and *annealed*; therefore, the search space is effectively reduced to enable efficient sampling within given computational resources. In SA, the system temperature provides thermal fluctuations, but the equipartition theorem prohibits the selective heating of specific degrees of freedom. SQ is conceptually similar to potential energy smoothing [18] in that a smoothly deformed potential energy surface, smoothed according to smoothing schedule, is explored. However in SQ, the locations of all the local minima and the overall topology of the potential energy surface are maintained during Q-annealing.

There is an important unresolved question in implementing SQ. A cooling scheme for q should be provided for SQ to work in practice. Direct scaling of q from its initial value to 1.0 with a simple function in simulated Q-annealing can be no more efficient than direct scaling of T with a simple function in SA [19]. There could be a (near) optimal Q-annealing protocol which can be supported by rigorous mathematics. One candidate may be the Q-annealing with the inverse function of the visiting temperature introduced in Ref 8. The target temperature, in this case, can be operationally defined in terms of q with the help of Eq. 3. In this work, the direct scaling of q with a linear and sigmoidal function is to be tested. This will provide the initial reference, possibly the lower bound, for the performance of SQ. At least, this direct scaling of q is conceptually similar to the direct scaling of energy [20].

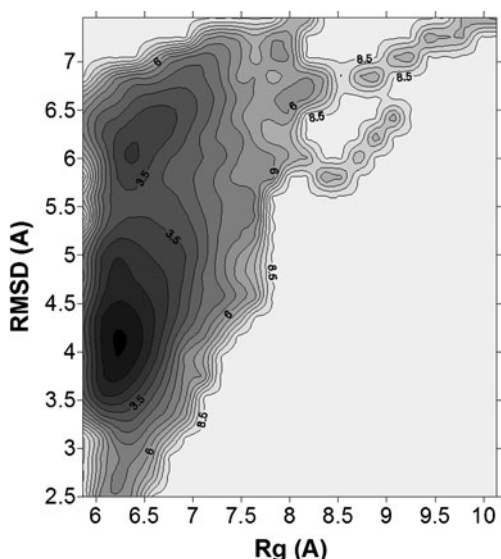


Fig. 1 Free energy surface of 1AQG as a function of R_g and $RMSD$ at $T=300$ K generated from the REMD simulation. Here R_g is mass-weighted radius of gyration including all atoms, and $RMSD$ is root-mean-squared distance with respect to 1AQG on the PDB database

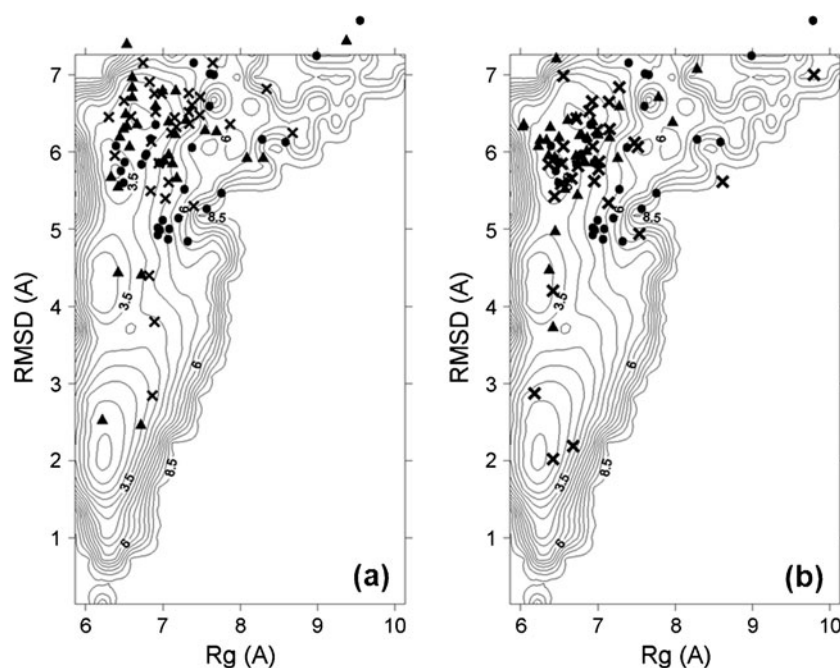
Simulation details

Simulated Q-annealing (SQ), or the molecular dynamics version of the generalized simulated annealing in the q -space, was implemented in the ANNEAL program of TINKER [21]. Here, q is the parameter of the Tsallis statistics, found in Eqs. 1 and 2. Only solvation, electrostatic, van der Waals, and torsional energy were considered in the effective

potential. Both SA and SQ were applied to 1AQG [22], a synthetic 11-residue protein segment (IKENLKDCGLF) derived from the heterotrimeric GTP-binding protein transducin α subunit C-terminal $GT\alpha$ (340–350). For comparative reasons, other simulation conditions were kept identical. All simulations were performed under a CHARMM27 force field. A velocity version of Verlet integrator with RATTLE was used. The time step was 2 fs. A generalized Born solvation model [23–25] was adopted to take into account solvation effects.

The starting configurations were identical for both of the SA and SQ simulations. A fully extended chain of 1AQG was generated, and then molecular dynamics with the initial configuration was performed during 6 ps under the CHARMM27 force field. Among the trajectories, 31 configurations were chosen regularly in the perspective of time interval. Then, SA and SQ were applied to the same configuration set of 31 members. Not a single step was dedicated to relaxation before or after the SA and SQ; hence identical initial configurations were launched into different annealing schemes. For the cooling schedule in T or q , both linear and sigmoidal scalings were tested. For SA, the initial temperature was set to 1000 K, and the final temperature to 300 K. In the ANNEAL program performing SA in temperature space, an annealing protocol was implemented by allowing smooth changes between the initial and final temperatures through a coupling to an external bath [21]. For SQ, the initial q -value was 1.0007 with a final q -value of 1.0000, while the temperature was kept constant (at 300 K) using a Berendsen thermostat. By

Fig. 2 Coordinates of the initial and final configurations are marked on the reconstructed free energy landscape. Filled circle (●) stands for the initial configurations, filled triangle (▲) for the final configurations of SQ, and cross (x) for the final configurations of SA, respectively. Linea cooling modes are illustrated in (a), sigmoid modes in (b)



Eq. 3 the effective temperature from the above value of q in SQ is estimated approximately to be 582 K. The effective temperature is not to be distributed among all the degrees of freedom; hence the smaller effective temperature in SQ is selected. Cooling steps were 1,000,000 in both methods. An additional parameter ε required in SQ was set to 800.0.

Results and discussion

The reported conformation ensemble on the PDB database is based on the experimental results of NOESY spectra, minimized as bound to photoexcited rhodopsin (covalently attached to *ILE-339*), under a CHARMM22 force field [22]. Since conformations of isolated 1AQG sampled under the CHARMM27 force field at 300 K are the main interest here, the representative free energy minimum structure as a reference for this condition should be prepared in a separate way. For that purpose, classical replica exchange molecular dynamics (REMD) simulations were performed as implemented in TINKER [21]. Trajectories were sampled up to 20 ns, with eight replicas (at 300.0 K, 327.0 K, 358.0 K, 392.0 K, 431.0 K, 473.0 K, 520.0 K, and 578.0 K). The overall exchange ratio was kept to around 22 %. The time step was 2 fs, and trajectories were generated by a velocity version of the Verlet algorithm with RATTLE. The temperature was kept constant with a Berendsen thermostat. For the consideration of solvation effects, a generalized Born solvation model [23–25] implemented in TINKER was applied.

Figure 1 shows the free energy surface as a function of R_g and $RMSD$ at $T=300$ K generated from the REMD simulation. The free energy surface was constructed from the resulting population distribution using $\Delta F(R_g, RMSD) = -RT \ln P(R_g, RMSD)$ where R_g is the mass-weighted radius of gyration including all atoms, and $RMSD$ is the root-mean-squared distance with respect to 1AQG on the PDB database. The free energy minimum structure based on this free energy surface is to be considered as the reference native structure compatible with the simulation conditions described above. To avoid high degeneracy in structures due to the low resolution of the free energy grids, all the atoms were included in the $RMSD$ calculation. $RMSD$ were calculated with the quaternion-based algorithm [26]. Usually trajectories of initial stage of simulations are not to be taken into consideration in free energy surface construction, but they are included in the present case in order to compare them with annealing trajectories. The overall topology of the free energy surface is funnel-like with the two deep minimum regions. The global minimum point on the free energy surface is $(R_g, RMSD) = (6.22, 4.15)$, and the potential energy minimum structure corresponds to the point $(6.07, 4.39)$. A representative free energy minimum struc-

ture, chosen from the free energy minimum ensemble by its minimum potential energy, is stable by approximately $0.6 \text{ kcal mol}^{-1}$ in the free energy landscape. In further analysis on SA and SQ, $RMSD$'s were calculated with respect to the free energy minimum structure found through REMD simulation and the free energy surface was also reconstructed.

With the configurations regularly selected among the early 31 trajectories of the canonical MD simulation as initial configurations, we performed SA and SQ both with linear and sigmoid cooling schemes, respectively. Figure 2 shows the positions of initial and final configurations on the free energy landscape for each simulation. It is noted that the performance of SQ is comparable with that of SA with linear cooling mode. However, in sigmoid mode, the performance of SQ is not appealing. One of the reasons for this discrepancy between SQ in linear and sigmoid

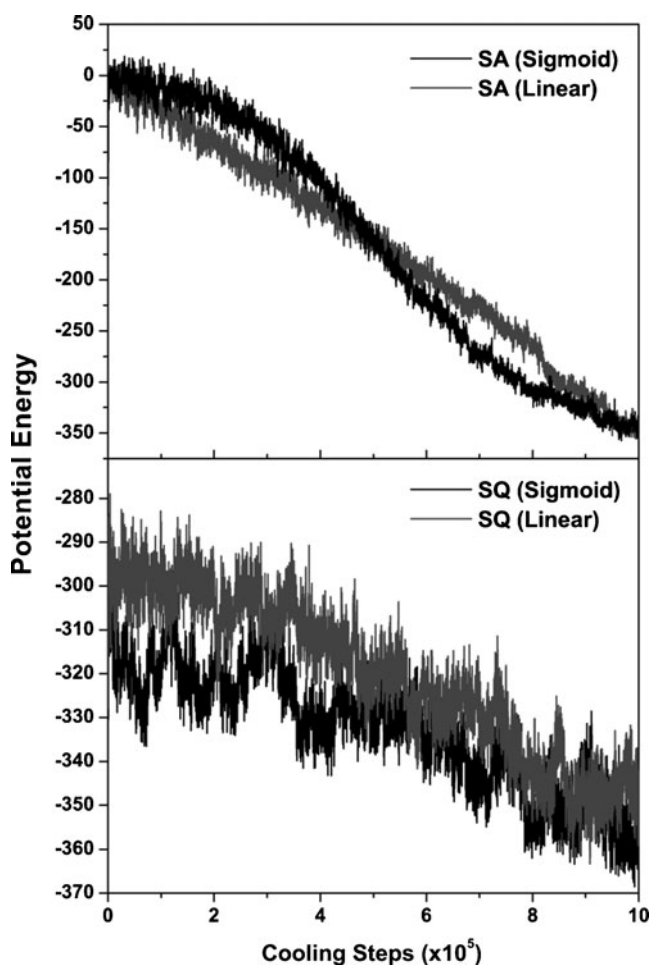


Fig. 3 Time profiles of potential energies for (a) the ordinary simulated annealing (SA) and (b) the simulated Q-annealing (SQ) simulations for 1AQG. Sigmoidal (black line) and linear (gray line) scaling schemes for the values of the temperature (SA) and the q parameter (SQ) are used for the annealing simulations. The units are arbitrary with the scales of the potential energy profiles to be identical for the two cases

modes can be attributed to the fact that the sigmoid mode spends more time in sampling with small q . When q is small, it is more likely that the equation of motion is dominated by ε . Actually if we reduce the value of ε by half, the final configurations reflect the free energy landscape better. It can be argued that the interplay between q and ε needs to be understood in order to fully exploit the potential of the SQ.

For further analysis, we focus on the annealing with the fully extended initial configuration. Figure 3 shows the potential energy profile for SA and SQ on the fully extended initial configuration. It should be noted that no relaxation stage was provided before SA and SQ. Annealing from $T=1000.0$ K and $q=1.0007$ was applied to the initial configuration which was generated at $T=300.0$ K and $q=1$. In the case of SA, the potential energy followed the temperature cooling profile. On the contrary, the potential energy for both scaling scheme in SQ showed a general trend of decreasing with large fluctuations. As discussed before, ε and q values are coupled to the effective

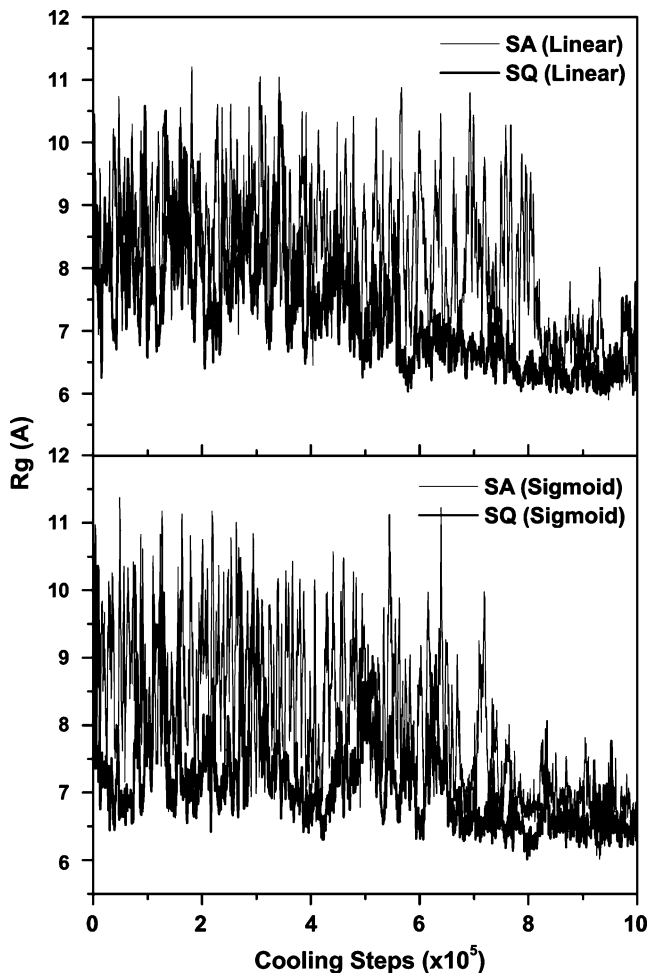


Fig. 4 Time profiles of R_g (in units Å) for SA (black line) and SQ (gray line) simulations for 1AQG with (a) linear and (b) sigmoidal scaling schemes for the annealing parameters

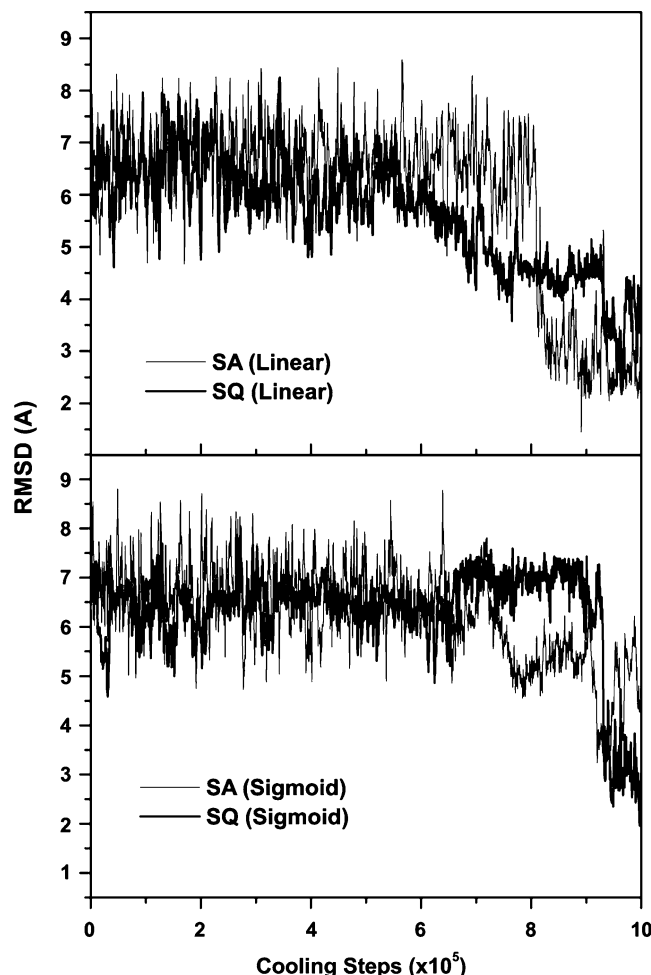


Fig. 5 Time profiles of $RMSD$ (in units Å) for SA (black line) and SQ (gray line) simulations for 1AQG with (a) linear and (b) sigmoidal scaling schemes for the annealing parameters

temperature by Eq. 3. Therefore, the scaling of q values may be related to that of temperature. Note that the potential energy for SQ in Fig. 3 is not equal to the effective potential given by a Tsallis transformation (Eq. 2). It can be argued that simple q -scaling does not ensure the



Fig. 6 Schematic diagram illustrating conformational searching performed by SA and SQ. In SA, conformations are sampled as the system energy is lowered according to the cooling scheme, as represented by the gray line on the left figure. On the contrary, conformational searching during SQ is done with small energy variations and with the changing barrier height according to the variations of q parameters, as illustrated in the right figure

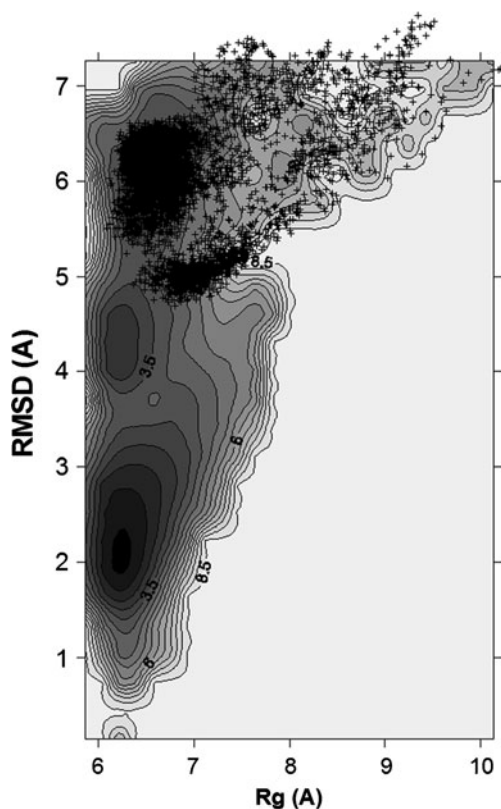
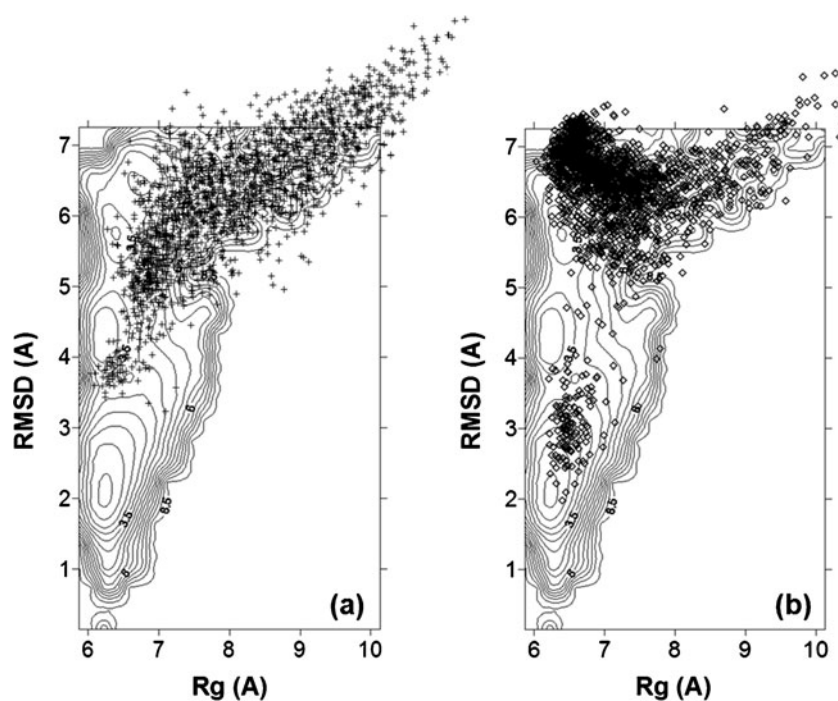


Fig. 7 Conformations visited during classical molecular dynamics simulation at 300.0 K for 1AQG are projected on the free energy surface (marked with +). The free energy surface is reconstructed with new *RMSD* values calculated with respect to the free energy minima structure (see Fig. 1)

Fig. 8 Conformations visited during (a) SA (marked with +) and (b) SQ (marked with \diamond) annealing simulations with sigmoidal scaling schemes for 1AQG are projected on the free energy surface

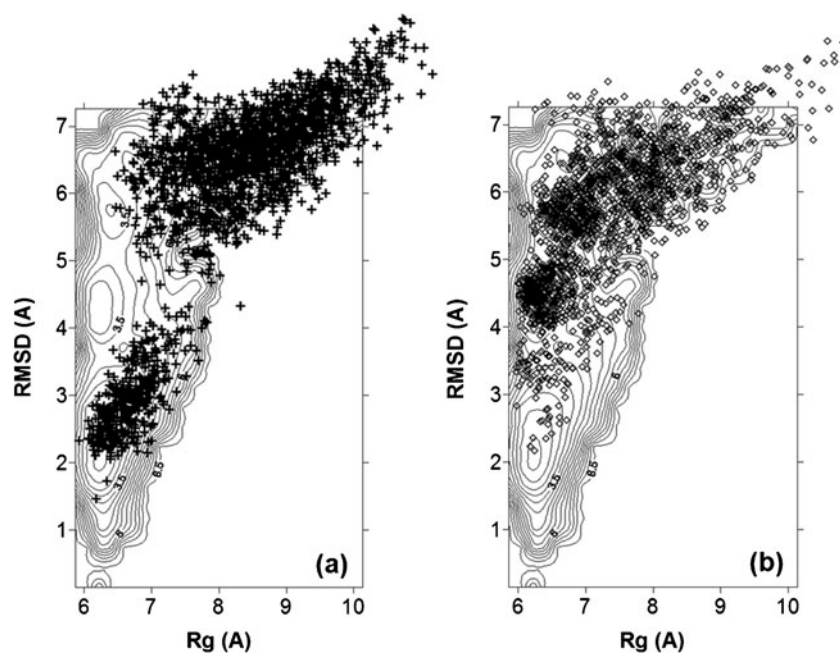


full coupling of scaled potential and system response and the adopted scaling schemes for SQ are not optimal annealing protocols.

Time profiles of R_g and *RMSD* for SA and SQ are shown in Figs. 4 and 5. At the end of the annealing processes, both R_g and *RMSD* are reduced as expected. During SA, they showed relatively large fluctuations without distinct features. On the contrary, throughout the Q-annealing, changes in R_g and *RMSD* seemed to show patterns with less fluctuation. It may be argued that R_g of SQ in the sigmoidal cooling scheme shows an oscillating pattern. This behavior can be attributed to the different mechanism of sampling enhancement for the two annealing methods. In SA, a configuration initially wanders about up in the potential energy surface due to its high kinetic energy then slowly goes down according to the cooling protocol. As a result, topological details near the potential energy minima cannot be explored until the system is cooled close to the lower temperature. On the other hand, in SQ, initial configuration is changing according to the gradient of the potential energy surface at an almost fixed height from the global minima, while the energy barrier changes as q varies. Therefore, topology of the potential energy surface is reflected throughout the annealing process. This conceptual difference between the two annealing methods is illustrated in Fig. 6.

As a reference data, trajectories of classical molecular dynamics simulation at 300.0 K are projected onto the free energy surface (Fig. 7). The free energy surface is reconstructed with new *RMSD* values calculated with

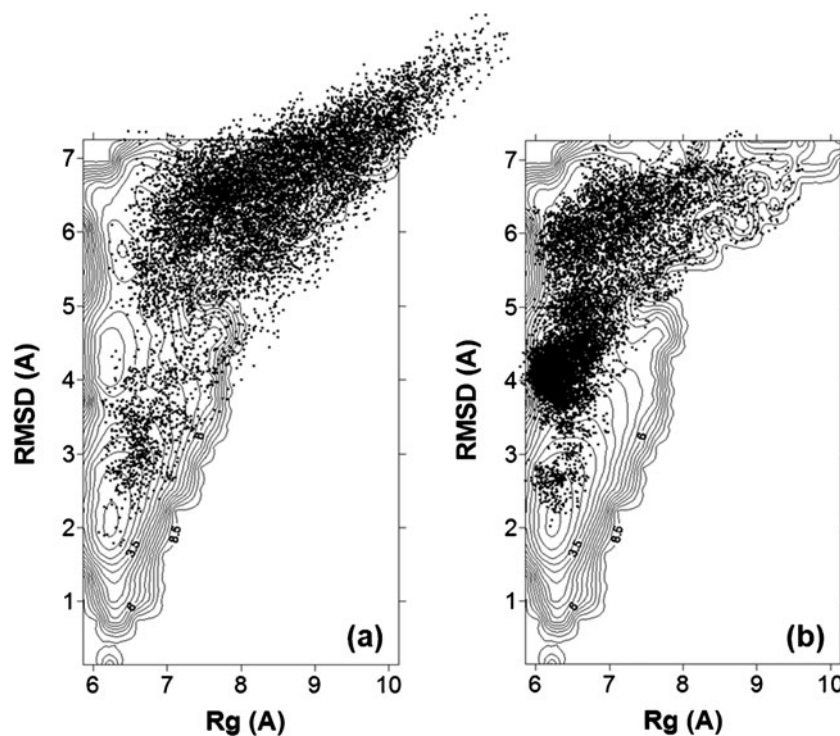
Fig. 9 Conformations visited during (a) SA (marked with +) and (b) SQ (marked with \diamond) annealing simulations with linear scaling schemes for 1AQG are projected on the free energy surface



respect to the free energy minima structure. It clearly shows that MD trajectories are trapped in local minima. Conformations visited during the cooling stage for the two annealing methods are projected onto the free energy landscape in Figs. 8 and 9. As pointed out above, ensemble of conformations sampled through SQ reflect the topology of the potential energy surface better than those through

SA. With linear scaling, conformations sampled with SA are rather uniformly distributed on the free energy surface and failed to cover sufficiently either of the free energy minimum regions. For SQ with the same linear scaling, the sampled conformations are concentrated around the two minimum regions including the global minimum. With the sigmoidal scaling, SA was able to sample conformations

Fig. 10 Conformations visited during (a) SA and (b) SQ annealing simulations with linear scaling schemes are projected on the free energy surface. The annealing trajectories are for a cluster of positions of the final configurations around $(R_g, RMSD) = (6.7, 2.5)$ as presented in Fig. 2



around the global minimum region. Even though the coverage of the global minimum region by the conformations from SQ in this case may be somewhat limited, one can argue that a wider region of the relevant free energy surface is explored by SQ.

In order to compare the performances of SA and SQ with the linear scaling, we also analyzed the trajectories for the simulations presented in Fig. 2. We selected a cluster of positions for the final configurations around $(R_g, RMSD) = (6.7, 2.5)$, which is close to the global minimum. Conformations corresponding to the annealing trajectories for the cluster with SA and SQ are projected onto the free energy landscape in Fig. 10. Even though SA sampled the global minimum region of the free energy map reasonably, SQ clearly represents the whole free energy surface much better. These results illustrated that SQ can reflect the topology of the free energy surface effectively.

Conclusions

In conclusion, simulated Q-annealing (SQ) can navigate the potential energy surface efficiently with a simple scaling annealing scheme. Performance is at least comparable to conventional SA with a well-defined cooling protocol. It can be argued that conformations sampled by SQ can represent the funnel-like free energy surface better than the ordinary simulated annealing (SA) method. SA has been incorporated with Monte Carlo Markov chain simulations, known as replica exchange methods. When SQ is combined with coupled Markov chain simulations, a more efficient conformation search method can be established. SQ works well with only a few selected degrees of freedom, which can also contribute to efficient conformational searching. This novel conformation search method, termed Q-REM, can be considered an embodiment of this idea within the framework of replica exchange molecular dynamics [27, 28].

Acknowledgments This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MEST) (No.2009-0079594 and No.2010-0001630).

References

- Pincus M (1970) A Monte Carlo method for the approximate solution of certain types of constrained optimization problems. *Oper Res* 18:1225–1228
- Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983) Optimization by simulated annealing. *Science* 220:671–680
- Wille LT (1986) Searching potential energy surfaces by simulated annealing. *Nature* 324:46–48
- Stützle T (2003) Simulated Annealing, Dynamic Local Search, GRASP, Iterated Greedy – an Overview. MN Summer School, Tenerife
- Hajek B (1988) Cooling schedules for optimal annealing. *Math Oper Res* 13:311–329
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell PAMI* 6:721–741
- Xiang Y, Gong XG (2000) Efficiency of generalized simulated annealing. *Phys Rev E* 62:4473–4476
- Tsallis C, Stariolo DA (1996) Generalized simulated annealing. *Physica A* 233:395–406
- Andricioaei I, Straub JE (1996) Generalized simulated annealing algorithms using Tsallis statistics: application to conformational optimization of a tetrapeptide. *Phys Rev E* 53:R3055–R3058
- Nayem A, Vila J, Scheraga HA (1991) A comparative study of the simulated-annealing and Monte Carlo-with-minimization approaches to the minimum-energy structures of polypeptides: [Met]-Enkephalin. *J Comput Chem* 12:594–605
- Wenzel W, Harmacher K (1999) Stochastic tunneling approach for global minimization of complex potential energy landscapes. *Phys Rev Lett* 82:3003–3007
- Hansmann UHE, Wille LT (2002) Global optimization by energy landscape paving. *Phys Rev Lett* 88:068105
- Schug A, Wenzel W, Hansmann UHE (2005) Energy landscape paving simulations of the trp-cage protein. *J Chem Phys* 122:194711
- Schug A, Herges T, Verma A, Lee KH, Wenzel W (2005) Comparison of stochastic optimization methods for all-atom folding of the Trp-cage protein. *Chem Phys Chem* 6:2640–2646
- Verma A, Schug A, Lee KH, Wenzel W (2006) Basin hopping simulations for all-atom protein folding. *J Chem Phys* 124:044515
- Northrup SH, McCammon JA (1980) Simulation methods for protein structure fluctuations. *Biopolymers* 19:1001–1016
- Andricioaei I, Straub JE (1997) On Monte Carlo and molecular dynamics methods inspired by Tsallis statistics: methodology, optimization, and application to atomic clusters. *J Chem Phys* 107:9117–9124
- Pappu RV, Hart RK, Ponder JW (1998) Analysis and application of potential energy smoothing and search methods for global optimization. *J Phys Chem B* 102:9725–9742
- Bounds DG (1987) New optimization methods from physics and biology. *Nature* 329:215–219
- Brünger AT (1991) Simulated annealing in crystallography. *Annu Rev Phys Chem* 42:197–223
- Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1983) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690
- Kisselev OG, Kao J, Ponder JW, Fann YC, Gautam N, Marshall GR (1998) Light-activated rhodopsin induces structural binding motif in G protein α subunit. *Proc Natl Acad Sci USA* 95:4270–4275
- Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* 112:6127–6129
- Qui D, Shenkin PS, Hollinger FP, Still WC (1997) The GB/SA continuum model for solvation. *J Phys Chem A* 101:3005–3014
- Onufriev A, Bashford D, Case DA (2000) Modification of the generalized Born model suitable for macromolecules. *J Phys Chem B* 104:3712–3720
- Coutsias EA, Seok C, Dill KA (2004) Using quaternions to calculate RMSD. *J Comput Chem* 25:1849–1857
- Jang S, Shin S, Pak Y (2003) Replica-exchange method using the generalized effective potential. *Phys Rev Lett* 91:058305
- Son WJ, Jang S, Pak Y, Shin S (2007) Folding simulations with novel conformational search method. *J Chem Phys* 126:104906

A model for the shuttle motions of puerarin and daidzin inside the cavity of β -cyclodextrin in aqueous acetic acid: insights from molecular dynamics simulations

Haiyang Zhang · Wei Feng · Cong Li · Yongqin Lv · Tianwei Tan

Received: 12 November 2010 / Accepted: 9 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract Acetic acid acts as one component of the mobile phase to influence separation of puerarin from daidzin when using β -cyclodextrin-substituted media. In this work considering an explicit acetic acid solution, host-guest complexes of β -cyclodextrin (β -CD) with puerarin and daidzin were investigated by molecular dynamics simulations. Computational results indicate different shuttle motions of puerarin and daidzin inside the cavity of β -CD. A model detailing the shuttle motion was constructed, and the relationships between shuttle depth and guest rotation angles, hydrogen bonds, and host-guest interaction energies were analyzed. The results can be used to explain the chromatographic retention mechanisms of puerarin and daidzin with β -CD, and to explore the complexity of host-guest interactions involving β -CD.

Keywords Molecular dynamics · β -cyclodextrin · *Pueraria* flavones · Complex · Acetic acid

Introduction

β -cyclodextrin (β -CD) has been the subject of much interest in the field of host-guest supramolecular complexes in recent years, because of its increasing applications in the pharmaceutical and food industries, and in materials, environmental protection, and separation science [1–7]. β -CD is shaped like a truncated cone containing seven glucose units; the secondary hydroxyl end is more open

than the primary end. The individual glucose units are held in a C-1 chair conformation, and are joined together by α -1,4 glycosidic linkages to form a cyclic structure. The interior of the cavity is lined by three rings of hydrogen atoms bonded to the C-5 glycosidic oxygen atoms, and hydrogen atoms at position C-3. The glycosidic oxygen bridges can produce a high electron density; hence, the inside of the cavity is relatively hydrophobic compared with water [8, 9]. With the existence of hydroxyl groups on the outside of the molecule, the external faces are hydrophilic. The structural properties of β -CD give rise to a remarkable capacity to form inclusion complexes with various molecules, especially with aromatic compounds.

β -CD is applied extensively in separation technology [8] because CD-complexation phenomena provide the procedure of choice for separation and extraction process. As a ligand coupled on agarose gels and spherical beads, β -CD has been employed for the isolation and purification of pueraria flavones in our laboratory [10–13]. These experiments showed that, when using β -CD-coupled medium as the stationary phase, acetic acid (HAc) as one component of the mobile phase effected the separation of the two substances puerarin and daidzin—two major isoflavones extracted from *Radix puerariae*. These two compounds have a similar structure and hence are difficult to isolate from each other. The molecular structures of β -CD, puerarin, and daidzin are shown in Fig. 1. Molecular dynamics (MD) studies on supramolecular complexes with cyclodextrins as host molecules have been published increasingly frequently in recent years [14–20], revealing MD simulation to be a useful and versatile tool for predicting the binding mode and exploring the fundamental principles that govern the binding of a guest molecule to its host at a molecular level. In previous MD studies [21], we simulated the formation process of inclusion complexes of

H. Zhang · W. Feng · C. Li · Y. Lv · T. Tan (✉)
Beijing Key Lab of Bioprocess, College of Life Science and Technology, Beijing University of Chemical Technology,
Beijing 100029, China
e-mail: twtan@mail.buct.edu.cn

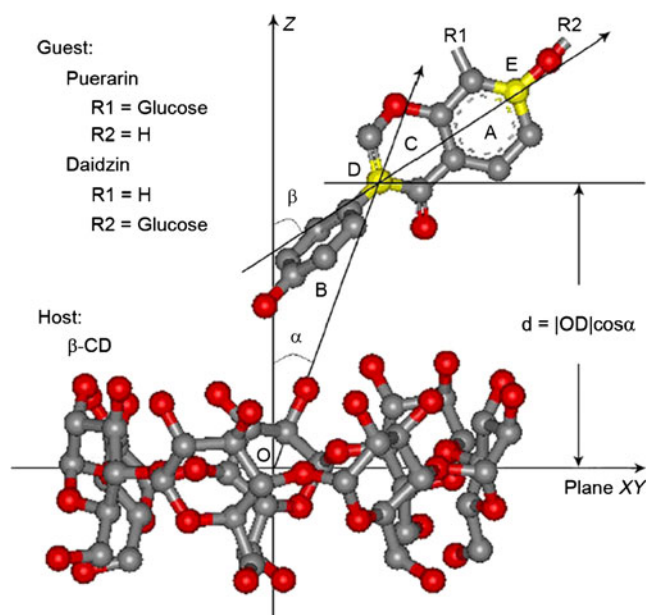


Fig. 1 Ball and stick models of guest (puerarin and daidzin) and host [β -cyclodextrin (β -CD)] molecules and definition of the shuttle depth d and rotation angle β depicting the relative position of a guest to its host. A , B , and C denote relevant ring groups, and D and E (yellow) are carbon atoms

puerarin and daidzin with β -CD in water and discussed the effects of agarose media and acetic acid on β -CD conformational changes induced by its guest molecule. In this report, we focus on shuttle motions of puerarin and daidzin inside the cavity of β -CD in HAC/water solvents through MD simulations, which serves as a further study on supramolecular interactions of puerarin and daidzin with β -CD. This work provides more insight into the mechanism behind such separations.

Computational methods

The software used for MD simulations was the GROMACS package (version 3.3) [22, 23] with the GROMOS-96 force field. Molecular topology parameters of β -CD, puerarin, daidzin, and HAC were constructed using the Dundee PRODRG Server [24]. Initial inclusion configurations of β -CD with puerarin and daidzin were obtained from previous MD studies. The degree of ionization of HAC was set to zero. All systems were simulated in a truncated octahedron box with an image distance of 5 nm in the NPT ensemble. The simulation box contained a single host-guest complex and around 3,000 solvent molecules, including HAC and simple point charge water molecules, corresponding to a density of about 1 g/mL. After equilibration via NPT, the box size was almost unchanged. Details of the simulation protocol have been presented in 21. All simulations were calculated with a time step of

0.002 ps, and atomic coordinates were saved to disk every 500 steps for later analysis. The overall simulation time of a single trajectory was 10 ns. The concentration of HAC/water solvents discussed in this work were 0%, 0.1%, 0.5%, 5%, 10%, and 20% (v/v), which were simulated before use to reach equilibration. Data processing was carried out using the programs within the GROMACS package and Discovery Studio Visualizer.

Results and discussion

In the initial state, the C ring of guest molecules (puerarin or daidzin) is inside the cavity of β -CD with the glucose unit close to the more open secondary end. Calculation results reveal that, due to interactions between the host, guest, and solvent molecules, both puerarin and daidzin can shuttle inside the cavity of β -CD in the HAC/water solvent. The shuttle motions of guest molecules inside the cavity cause changes of the relative position between the host and guest molecules, and hence changes in host-guest interactions.

As shown in Fig. 1, two parameters, a shuttle depth (d) and a rotation angle (β), are used to indicate the position of the guest molecule. As seven glycosidic oxygen atoms of β -CD are located almost in one plane (designated as plane XY), O is set to be the centroid of the seven oxygen atoms, and the axis Z is defined as the normal vector of the plane XY with the narrow end of β -CD below the plane XY. The carbon atoms D and E (yellow balls in Fig. 1) belong to the C and A ring of the guest molecules, respectively. The shuttle depth is calculated by $d = |OD| \cos\alpha$, where α is the angle between the Z axis and the vector from point O to point D, and $|OD|$ is the distance between the points O and D. The rotation of the guest molecule around the axis Z is characterized by the angle β between the Z axis and the vector starting from point D, terminating at point E.

For the complex β -CD/puerarin in aqueous solution containing 20% HAC, the distance d frequently vibrates between the negative and positive values during the first 3-ns simulation, implying a shuttle behavior of puerarin inside the β -CD cavity. After 4 ns, puerarin gets out of the cavity via the secondary rim of β -CD. A similar phenomena was observed for the system of β -CD/daidzin in 20% HAC. Thus, the concentration of 20% HAC is not discussed further below.

Figure 2 shows the profile of the angle α as a function of the shuttle depth in HAC/water solvents. For $d=0$ nm and $\alpha=90^\circ$, the carbon atom (point D) of the C ring is on plane XY. When $\alpha>90^\circ$, d is negative, and point D is below the plane XY. The more negative the shuttle depth, the deeper the C ring of the guest molecules is embedded within the cavity with respect to the secondary end of β -CD. For the

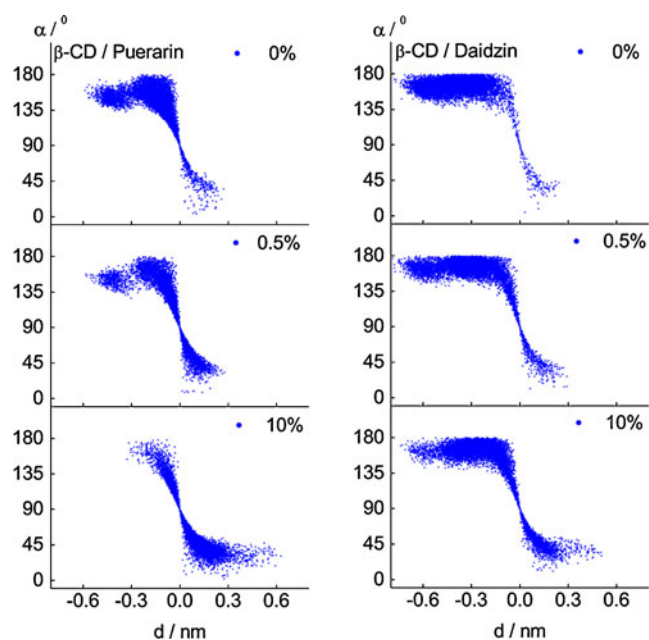


Fig. 2 Angle α vs the distance d for the complex β -CD/puerarin (*left*) and β -CD/daidzin (*right*) in acetic acid (HAc)/water solvents

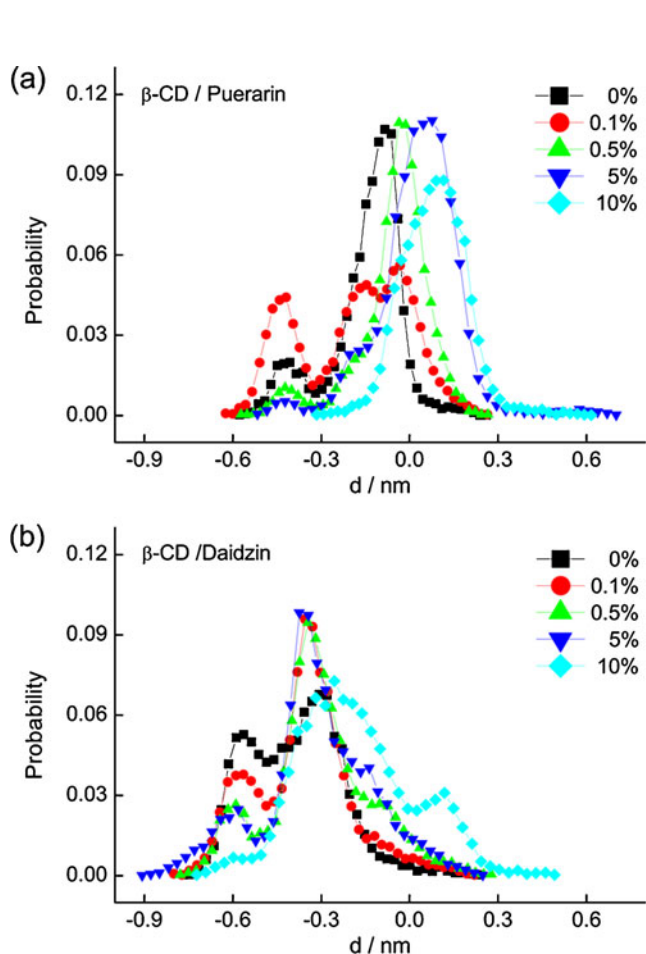


Fig. 3 Frequency distributions of the distance d for the complex β -CD/puerarin (a) and β -CD/daidzin (b) in HAc/water solvents

systems of β -CD/puerarin and β -CD/daidzin with 0% HAc, the angle α and distance d are in the range of $d < 0$ nm and $\alpha > 90^\circ$, respectively; with 10% HAc, for β -CD/puerarin, the angle α and distance d are shifted to the range of $d > 0$ nm and $\alpha < 90^\circ$; for β -CD/daidzin, although the angle α and distance d are partially shifted to the range of $d > 0$ nm and $\alpha < 90^\circ$, they are mainly in the range of $d < 0$ nm and $\alpha > 90^\circ$. Frequency distributions of the distance d in HAc/water solvents are shown in Fig. 3. As the concentration of HAc increases from 0% to 10%, the distance d is shifted to bigger values, and the guest gets gradually further away from the cavity of β -CD, indicating the elution effect of acetic acid on the separation process. The shifts for puerarin are more sensitive to the concentration of HAc than those of daidzin. Compared with puerarin, daidzin behaves more like a linear molecule and hence daidzin is more deeply embedded within the cavity of β -CD with respect to the secondary end, as shown in Fig. 3. This indicates that puerarin is desorbed more easily than daidzin from the cavity of β -CD in the HAc/water mobile phase, which is in

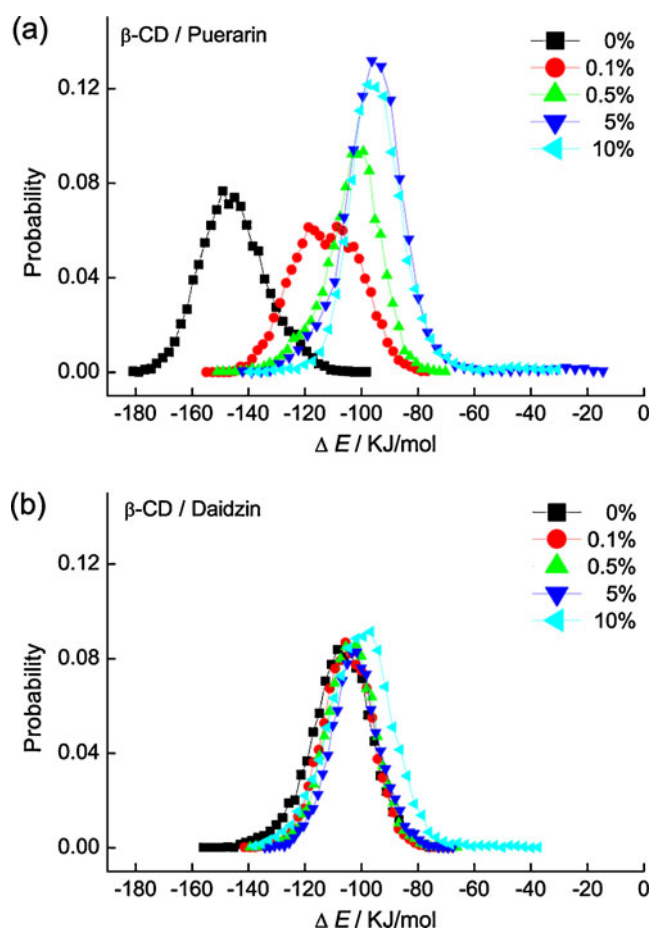


Fig. 4 Frequency distributions of host-guest interaction energy for the complex β -CD/puerarin (a) and β -CD/daidzin (b) in HAc/water solvents

accordance with experimental chromatograms where puerarin was first eluted [11].

Shuttle motions of puerarin and daidzin result in changes in host-guest interaction energy. The interaction energy between β -CD and its guest can be defined as the difference between the energy of the inclusion complex [$E(\beta\text{-CD}^*\text{guest})$] and the energy of the isolated host [$E(\beta\text{-CD})$] and guest [$E(\text{guest})$] in the same simulation box [25], as expressed by Eq. (1)

$$\Delta E = E_{\text{complex}} - E_{\text{host}} - E_{\text{guest}} \quad (1)$$

Regarding the β -CD-based complex, this expression may account, for example, for the unlike solvation of the host; for the strain energy, in particular of the macrocycle such as cyclodextrins; and for the hydrogen bonds of the water molecules clustered in the cavity and released upon inclusion, as noted by Raffaini et al. [26]. Frequency distributions of interaction energies between host and guest are presented in Fig. 4. In water (0% HAC), the interaction energy between

puerarin and β -CD is more negative than that between daidzin and β -CD, because the structure-induced effects of puerarin on the structure of β -CD in water are larger than those of daidzin [21]. Structural changes of β -CD can be reflected in the root mean square (RMS) deviations in structure that are calculated using the GROMACS tool “g_rms” by least-square fitting each structure of β -CD from a trajectory to the reference structure at time $t=0$. Larger RMS deviation means larger conformational fluctuation. For isolated β -CD in water, conformation changes are associated with a small fluctuation (the RMS deviation is equilibrated at 0.11 nm). However, the RMS deviation of β -CD for the complex β -CD/puerarin in water increases to 0.17 nm, due to the induced fit by puerarin. When HAC is added, the RMS deviation of β -CD from its initial configuration amounts to about 0.11 nm (data not shown), indicating that the structure of β -CD can be maintained well. With increasing concentrations of HAC, host-guest interaction energies decrease, and the guest molecule can escape from the cavity of β -CD at higher concentrations of HAC. We can see that, regarding

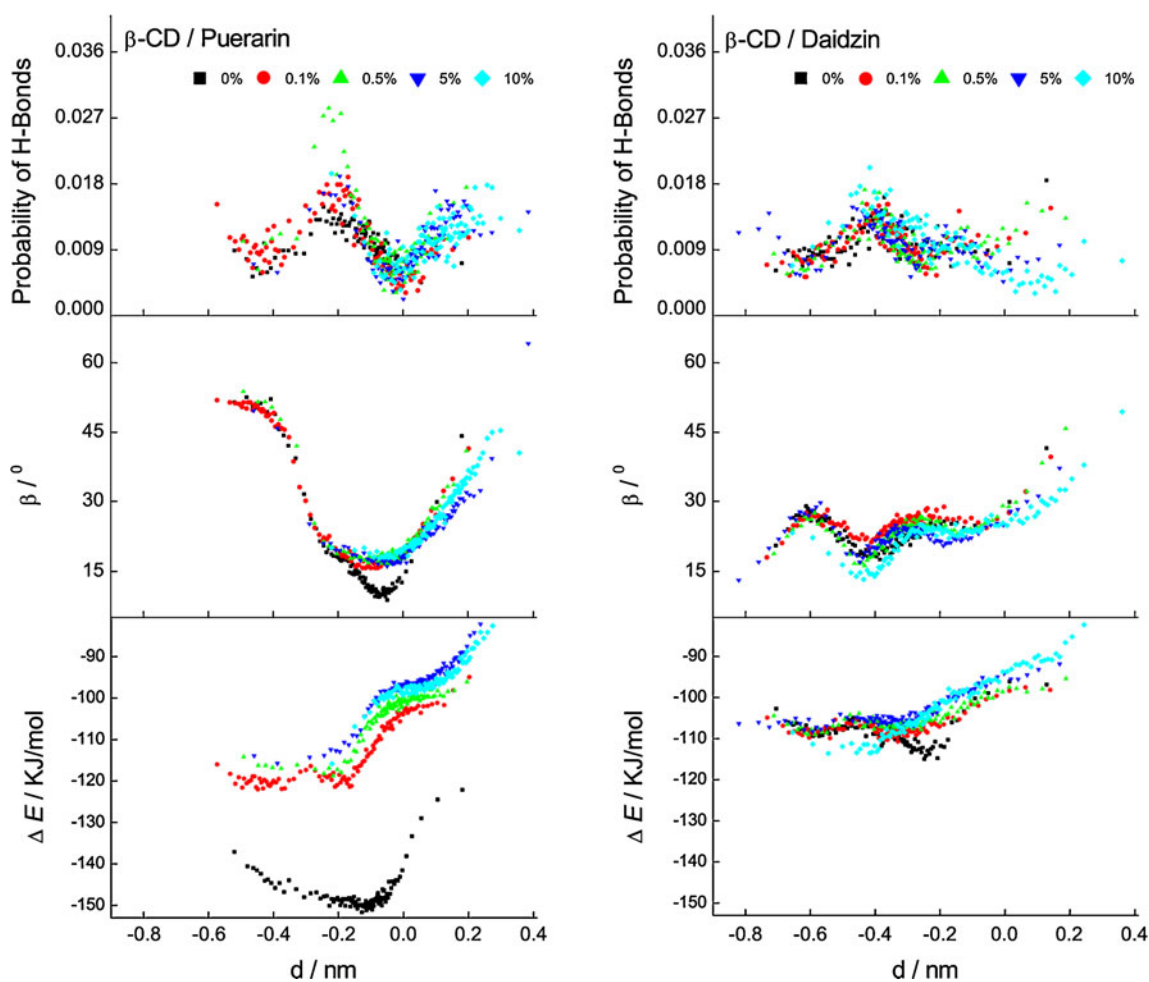


Fig. 5 Frequency distribution of the number of H-bonds between host and guest (*top*), angle β (*middle*), and host-guest interaction energy (*bottom*) vs the shuttle depth d for β -CD/puerarin (*left*) and β -CD/daidzin (*right*)

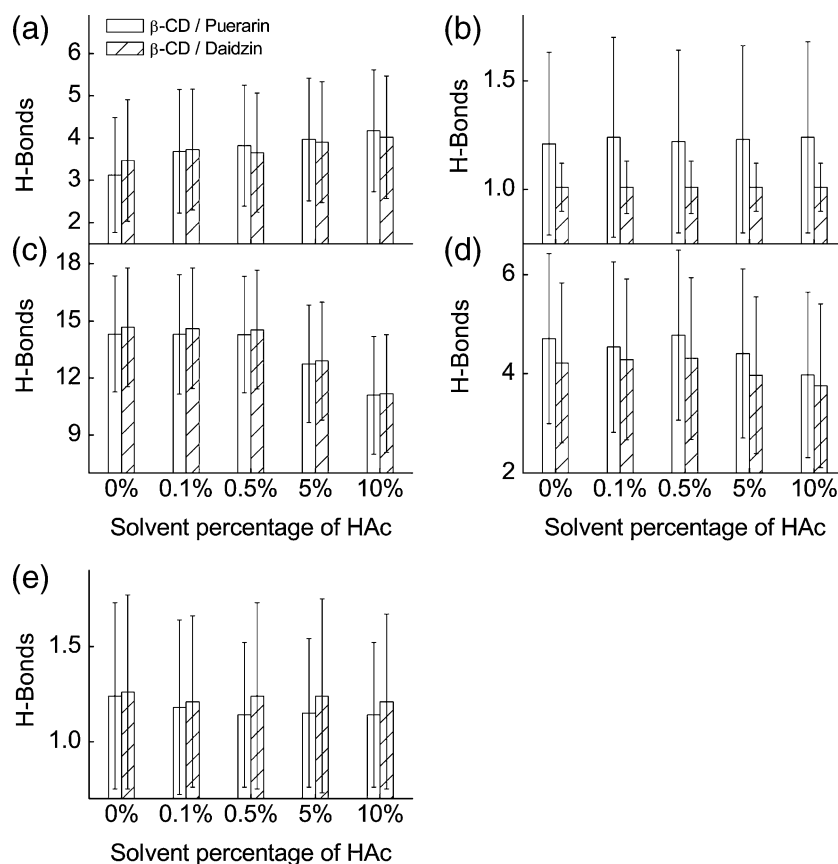
changes in host–guest interaction energy, puerarin is more sensitive to HAc than daidzin.

Relationships between the shuttle depth d and frequency distributions of the number of host–guest H-bonds, and the angle β and the host–guest interaction energy (ΔE) are shown in Fig. 5. As daidzin can insert more deeply into the cavity of β -CD than puerarin, the rotational movement of daidzin around the axis Z of β -CD is restricted. As a result, the amplitude of vibration of the angle β for daidzin is smaller than that for puerarin. For convenience, reference distances $d_{0.01}$ and $d_{-0.35}$, around which the data points (red, blue, green, and cyan) are much denser, are defined. As can be seen, the reference distances d are around 0.01 nm and -0.35 nm for β -CD/puerarin and β -CD/daidzin, respectively. The host–guest H-bonds are closely related to the relative position of guest to host. For β -CD/puerarin, when puerarin is positioned at distance d away from distance $d_{0.01}$, angle β increases and there are more hydrogen bonding interactions between β -CD and puerarin. For β -CD/daidzin, the situation is different. When daidzin is positioned with distance d away from distance $d_{-0.35}$, there are fewer hydrogen bonding interactions between β -CD and daidzin. Regarding the interaction energy for both systems in the solvent of HAc/water, when the guests are positioned with distance d larger than $d_{0.01}$ or $d_{-0.35}$, the

interaction energy decreases. The effect of HAc concentration can be observed with the data points in black, red, green, blue, and cyan in Fig. 5.

Detailed hydrogen bonding interactions between the host, guest, and solvents are analyzed and depicted in Fig. 6. As can be seen in Figs. 6a,c–e, the huge error bars compared to the average values preclude any clear-cut conclusion. For β -CD/puerarin and β -CD/daidzin, there are no statistically significant differences in the intramolecular H-bonds of β -CD (Fig. 6a) or in the intermolecular H-bonds between β -CD and HAc/water solvents (Fig. 6c), either between guest and solvents (Fig. 6d) or between host and guest (Fig. 6e). The average number of intramolecular H-bonds of puerarin is greater than that of daidzin (Fig. 6b). Considering groups connected to the A ring, puerarin has five hydrogen-bond donors and six acceptors, while daidzin has four donors and six acceptors. Thus, puerarin favors the formation of more intra- and intermolecular H-bonds. About 72% of intramolecular H-bonds for puerarin are formed between the glucose unit and the hydroxyl group connected to the A ring. We conclude that no single factor is actually primarily responsible for the differing stability of the complexes, but rather there are a few factors that can be conveniently summarized by the interaction energies reported below.

Fig. 6 The average number of intramolecular hydrogen bonds of host (a) and guest (b), and intermolecular hydrogen bonds between host and solvents (c), between guest and solvents (d), and between host and guest (e)



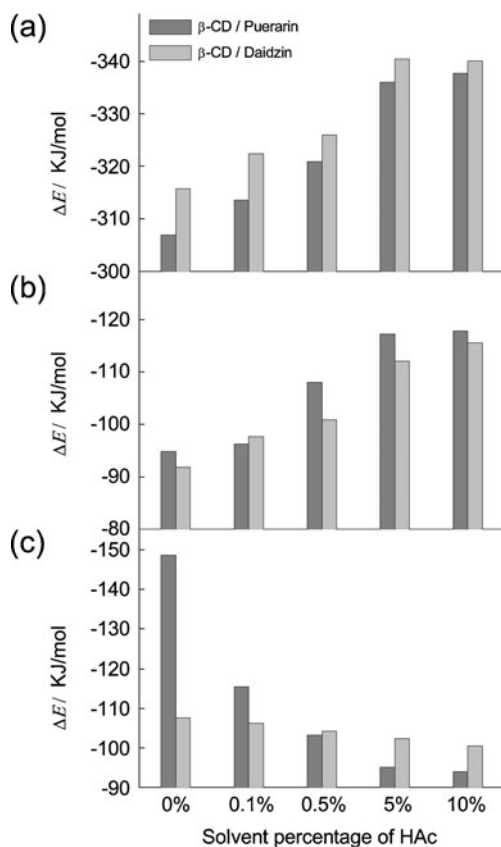


Fig. 7 Average interaction energies of 10-ns simulations between host and solvents (a), between guest and solvents (b), and between host and guest (c)

Figure 7 shows the average interaction energy of host–solvent, guest–solvent, and host–guest versus the concentration of HAC. The addition of acetic acid strengthens the interaction of host–solvent and guest–solvent, but makes the interaction between host and guest weaker. In higher concentrations of HAC (0.5%, 5%, and 10%), the interaction energy between β -CD and puerarin is weaker than with daidzin, which is consistent with chromatographic retention orders. Puerarin is eluted more easily in HAC/water than daidzin because of its stronger interaction with the solvent (Fig. 7b) and weaker interaction with the host (Fig. 7c).

Conclusions

Host–guest complexes of β -CD with puerarin and daidzin in a mixed solvent of acetic acid and water were investigated by MD simulations. Simulation results show the elution effect of acetic acid on the separation process and the shuttle motions of puerarin and daidzin inside the cavity of β -CD. A model for describing the relative position of a guest to its host was built to analyze shuttle motion related to shuttle depth, guest rotation angle around

the cavity axis, hydrogen bonds, and host–guest interaction energy. Daidzin can be embedded more deeply within the cavity with respect to the secondary rim of β -CD in HAC/water solvents than puerarin. Due to the hydroxyl group connected to the A ring, puerarin can form more intra- and intermolecular hydrogen bonds than daidzin, which results in stronger interactions with solvents, makes it easier to elute. Investigations of the shuttle behavior at a molecular level can be used to further explain experimental results and to explore the supramolecular complexity of interactions involving β -CD.

Acknowledgments This research was supported by the National Natural Science Foundation of China (20876011), 863 program (2006AA02Z245, 2007AA100404), 973 program (2007CB714304), Beijing Municipal Science & Technology Commission (Z09010300840901), and special funds of cooperation project of Beijing Municipal Education Commission.

References

- Miro G, Rondinone A, Nappi A, Ungaro F, Quaglia F, La Rotonda MI (2009) Modulation of release rate and barrier transport of diclofenac incorporated in hydrophilic matrices: role of cyclodextrins and implications in oral drug delivery. *Eur J Pharm Biopharm* 72:76–82
- Thi TD, Nauwelaerts K, Froeyen M, Baudemprez L, van Speybroeck M, Augustijns P, Annaert P, Martens J, van Humbeeck J, van den Mooter G (2010) Comparison of the complexation between methylprednisolone and different cyclodextrins in solution by $^1\text{H-NMR}$ and molecular modeling studies. *J Pharm Sci* 99:3863–3873
- Lu Z, Cheng B, Hu YL, Zhang YH, Zou GL (2009) Complexation of resveratrol with cyclodextrins: solubility and antioxidant activity. *Food Chem* 113:17–20
- Koontz JL, Marcy JE, O’Keefe SF, Duncan SE (2009) Cyclodextrin inclusion complex formation and solid-state characterization of the natural antioxidants α -tocopherol and quercetin. *J Agric Food Chem* 57:1162–1171
- van de Manakker F, Vermonden D, Nostrum CF, Hennink WE (2009) Cyclodextrin-based polymeric materials: synthesis, properties, and pharmaceutical/biomedical applications. *Biomacromolecules* 10:3157–3175
- Xu MY, Wu SZ, Zeng F, Yu CM (2010) Cyclodextrin supramolecular complex as a water-soluble ratiometric sensor for ferric ion sensing. *Langmuir* 26:4529–4534
- Si-Ahmed K, Tazerouti F, Badjah-Hadj-Ahmed AY, Aturki Z, D’Orazio G, Roccoa A, Fanali S (2010) Optical isomer separation of flavanones and flavanone glycosides by nano-liquid chromatography using a phenyl-carbamate-propyl- β -cyclodextrin chiral stationary phase. *J Chromatogr A* 1217:1175–1182
- Crini G, Morcellet M (2002) Synthesis and applications of adsorbents containing cyclodextrins. *J Sep Sci* 25:789–813
- Jullian C, Miranda S, Zapata-Torres G, Mendizábal F, Olea-Azar C (2007) Studies of inclusion complexes of natural and modified cyclodextrin with (+)catechin by NMR and molecular modeling. *Bioorg Med Chem* 15:3217–3224
- Wang MY, Kan WN, He XL, Tan TW, Janson JC (2005) Preparative purification of puerarin from pueraria flavones by oligo- β -cyclodextrin-Sepharose HP matrix. *J Liq Chromatogr Relat Technol* 28:1509–1518

11. He XL, Tan TW, Xu BZ, Janson JC (2004) Separation and purification of puerarin using β -cyclodextrin-coupled agarose gel media. *J Chromatogr A* 1022:77–82
12. Li R, Zhao R, Zhang HY, Li C, Feng D, Qin PY, Tan TW (2010) A novel medium poly(vinyl acetate-triallyl isocyanurate-divinylbenzene) coupled with oligo- β -cyclodextrin for the isolation of puerarin from pueraria flavones. *Chromatographia* 72:47–54
13. Yang L, Zhu Y, Tan TW, Janson JC (2007) Coupling oligo- β -cyclodextrin on polyacrylate beads media for separation of puerarin. *Process Biochem* 42:1075–1083
14. Lipkowitz KB (1998) Applications of computational chemistry to the study of cyclodextrins. *Chem Rev* 98:829–1874
15. Sun M, Liu XH, Yan LS, Luo GA, Zhao YF (2003) Molecular recognition between 4aS/R-galanthamine diastereoisomers and α -cyclodextrin. *J Mol Model* 9:419–422
16. Dodziuk H (2006) Cyclodextrins and their complexes. Wiley-VCH, Weinheim
17. Sellner B, Zifferer G, Kornherr A, Krois D, Brinker UH (2008) Molecular dynamics simulations of β -cyclodextrin–ziadamantane complexes in water. *J Phys Chem B* 112:710–714
18. El-Barghouthi MI, Jaime C, Al-sakhen NA, Issa AA, Abdoh AA, Ai Omari MM, Badwan AA, Zughul MB (2008) Molecular dynamics simulations and MM–PBSA calculations of the cyclodextrin inclusion complexes with 1-alkanols, *para*-substituted phenols and substituted imidazoles. *J Mol Struct THEOCHEM* 853:45–52
19. Figueiras A, Sarraguca JM, Pais AA, Carvalho RA, Veiga JF (2010) The role of L-arginine in inclusion complexes of omeprazole with cyclodextrins. *AAPS PharmSciTech* 11:233–240
20. Kokkinou A, Tsorteki F, Karpusas M, Papakyriakou A, Bethanis K, Mentzafos D (2010) Study of the inclusion of the (R)- and (S)-camphor enantiomers in α -cyclodextrin by X-ray crystallography and molecular dynamics. *Carbohydr Res* 345:1034–1040
21. Zhang HY, Feng W, Li C, Tan TW (2010) Investigation of the inclusions of puerarin and daidzin with β -cyclodextrin by molecular dynamics simulation. *J Phys Chem B* 114:4876–4883
22. van der Spoel D, Lindahl E, Hess B, Buuren AR, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers AL, Feenstra KA, Drunen R, Berendsen HJ (2005) Gromacs user manual (version 3.3). <http://www.gromacs.org/documentation/manual>
23. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AF, Berendsen HJC (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701–1708
24. Schuettelkopf AW, van Aalten DMF (2004) PRODRG—a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr D60*:1355–1363
25. Kollman PA, Massova I, Reyes C, Kuhn B, Huo SH, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TS (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33:889–897
26. Raffaini G, Ganazzoli F, Malpezzi L, Fuganti C, Fronza G, Panzeri W, Mele A (2009) Validating a strategy for molecular dynamics simulations of cyclodextrin inclusion complexes through single-crystal X-ray and NMR experimental data: a case study. *J Phys Chem B* 113:9110–9122

Fuzzy oil drop model to interpret the structure of antifreeze proteins and their mutants

Mateusz Banach · Katarzyna Prymula ·
Wiktor Jurkowski · Leszek Konieczny · Irena Roterman

Received: 18 December 2010 / Accepted: 7 March 2011 / Published online: 27 April 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Mutations in proteins introduce structural changes and influence biological activity: the specific effects depend on the location of the mutation. The simple method proposed in the present paper is based on a two-step model of in silico protein folding. The structure of the first intermediate is assumed to be determined solely by backbone conformation. The structure of the second one is assumed to be determined by the presence of a hydrophobic center. The comparable structural analysis of the set of mutants is performed to identify the mutant-induced structural changes. The changes of the hydrophobic core organization measured by the divergence entropy allows quantitative comparison estimating the relative structural changes upon mutation. The set of antifreeze proteins, which appeared to represent the hydrophobic core structure

accordant with “fuzzy oil drop” model was selected for analysis.

Keywords Antifreeze · Hydrophobicity · Intermediates · Mutants · Protein structure

Introduction

The mutation is a phenomenon observed in living cells. It is considered the main feature of evolution, modifying the structure of proteins, as well as their biological activity.

The modification of protein structure aimed at generating proteins with the desired biological function is currently a very popular issue.

The consequences of point mutations are reported in context of unfolding process [1, 2]. Temperature-jump induced transition state of ubiquitin in unfolding dynamic in WT and mutant forms of this downhill protein revealed the existence of the intermediate state in thermal unfolding of this protein [3–5]. The influence of the particular mutations on the unfolding process was examined for titin revealing that the I27 mutation demonstrates the opposite effect on protein stability in respect to Y9P [2]. The decreased pressure and temperature stability, the crystal structure of bovine pancreatic ribonuclease A variants V47A, V54A, V57A, I81A, I106A, and V108A was detected experimentally revealing the individual response to mutations [6].

The data base oriented on the collection of mutants form has been organized to integrate the structures changed upon mutation (<http://bioinformatics.eas.asu.edu/sprouts.html>) [7]. Linearly forced elastic network

M. Banach · K. Prymula · W. Jurkowski · I. Roterman (✉)
Department of Bioinformatics and Telemedicine, Medical
College, Jagiellonian University,
Lazarza 16,
31–530 Krakow, Poland
e-mail: myroterm@cyf-kr.edu.pl

M. Banach
Faculty of Physics, Astronomy and Applied Computer Science,
Jagiellonian University,
Reymonta 4,
30–059 Krakow, Poland

K. Prymula
Faculty of Chemistry, Jagiellonian University,
Ingardena 3,
30–060 Krakow, Poland

L. Konieczny
Medical College, Jagiellonian University,
Kopernika 7,
31–034 Krakow, Poland

model (LFENM) to characterize the mutational effects on structure appeared the general tool for the recognition of the observed pattern of structural divergence revealing that the normal modes dominate structural changes [8]. I-Mutant2.0 is a support vector machine (SVM)-based tool for the automatic prediction of protein stability changes upon single point mutations. I-Mutant2.0 can be used both as a classifier for predicting the sign of the protein stability change upon mutation and as a regression estimator for predicting the related $\Delta\Delta G$ values. The web interface allows the selection of a predictive mode that depends on the availability of the protein structure and/or sequence [9]. The cross-validated tests of a computational classifier, a support vector machine (SVM) was applied to classify the highly informative features of the best predictability of the functional annotation of the nucleotide sequence was presented in [10, 11]. The folding process influenced by mutation is the object of analysis [12, 13].

The set (the largest one found in PDB) of proteins representing different forms of the proteins belonging to antifreeze proteins is the object of analysis in this work. The attempt is undertaken to present the general model for quantitative and qualitative measurements of the consequences of the mutations. The structural changes are analyzed in respect to the model of folding process *in silico*. The two-step model treating the folding process as mediated by two intermediates (between unfolded state and the native one) is applied for comparable structural analysis [14, 15]. The structure of the first intermediate called early stage (ES) is assumed to be generated solely according to backbone conformation [16]. The traces of the ES intermediate characteristics is measured in the structures of proteins under consideration. The late stage (LS) intermediate is assumed to be generated as the effect of the influence of external force field of the hydrophobic character expressed by three-dimensional Gauss function representing the structure of hydrophobic core [17]. The accordance of the proteins structure with the hydrophobic core (the highest hydrophobicity density in the center of the protein and decreased with the increase of distance versus the center of the molecule body reaching values zero on the surface) and its changes are used to express the structural/functional changes. The biological activity seems to be affected by the changes of hydrophobic core structure.

Materials and methods

Two-step protein folding process

The protein folding process was recognized experimentally as multi-step process with unknown number of intermediates

[14, 15]. The model presented in this work assumes two-step process:



where : U – unfolded, ES – early stage, LS – late stage and N – native structural form.

Early stage model

This model assumes the dominant role of backbone, the conformation of which is expressed by two geometric parameters [15, 16]. The first one is the V-angle – the dihedral angle between two sequential peptide bond planes, the value of which is close to 0 deg for helical forms and close to 180 degs for extended and β -like structures. The second one, which seems to be determined by the first one, is the radius of curvature R of the polypeptide fragment (pentapeptide), which is small for helical structures and large for β -structural forms. The relation between these two parameters, which may apparently be expressed using a second degree polynomial,

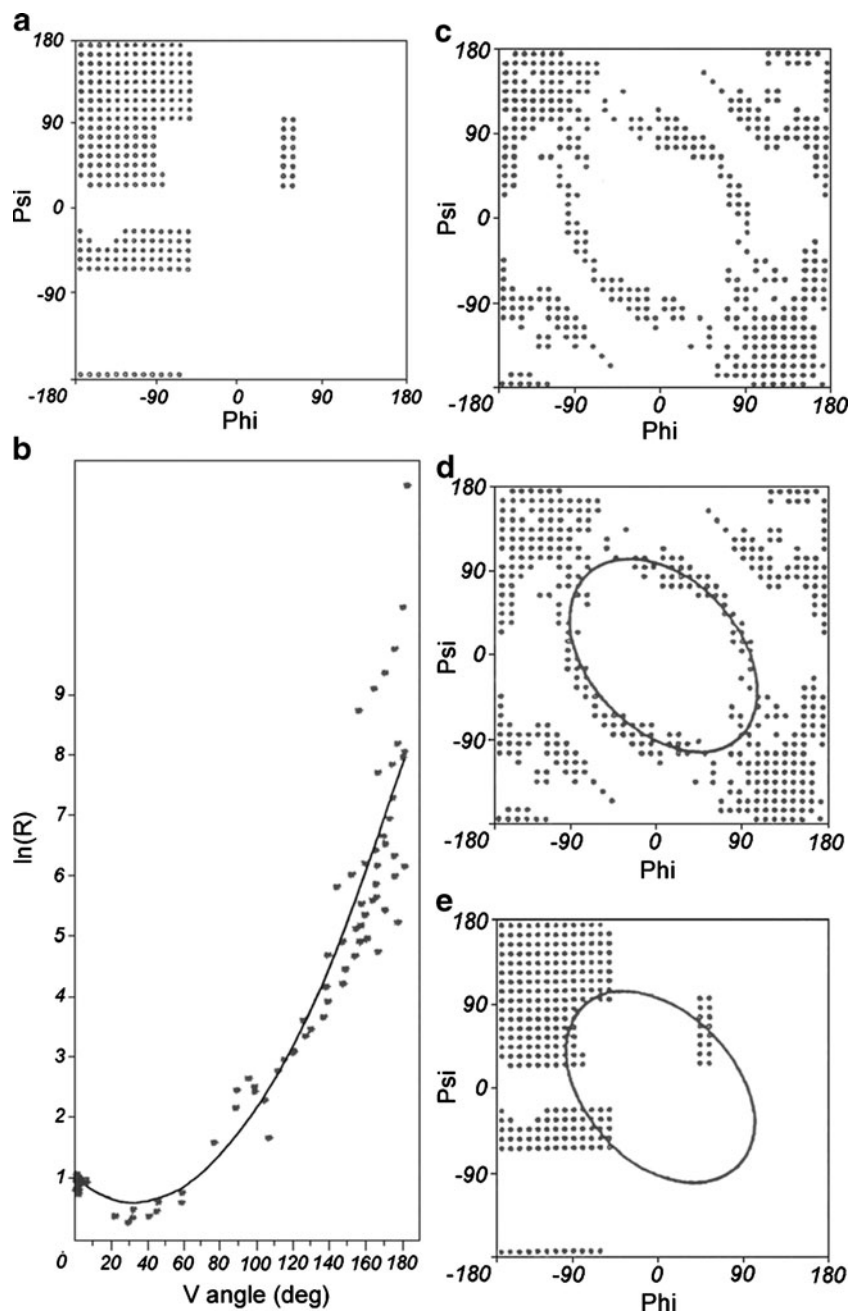
$$\ln(R) = 0.0003V^2 - 0.02009V + 0.848, \quad (1)$$

determines the optimal path on the Ramachandran map considered the complete conformational space. The elliptical path on the Phi-Psi map links the locations of all secondary structures. This path is assumed to represent the limited conformational sub-space available for the backbone in the ES step of the folding process. The agreement between the model and the protein is estimated by calculating the average distance ($D_{average}$) between the projected value of the radius of curvature and the one observed one for the appropriate V-angle value as it appears for particular residue in the polypeptide chain. The graphic interpretation of the ES model is given in Fig. 1.

Late stage model

The tertiary structure of the protein in the LS step of the protein folding process as assumed to be reached during the generation of the hydrophobic core with a simultaneous optimization of all other non-bonding interactions (electrostatic, vdW and torsional potential). The presence of an external force field is expressed via the three-dimensional Gauss function [17]. Model extends the original one introduced by Kauzman [18]. The force field simulates the hydrophobic core of the “fuzzy oil drop” model with the highest concentration of hydrophobicity in the center of the ellipsoid with its decrease depending on the distance from

Fig. 1 The ES model definition. (a) the Ramachandran map with low energy area distinguished (b) the relation between V-angle (dihedral angle between two sequential peptide bond planes) and R – radius of curvature (in logarithmic scale to avoid large values for β -structural forms) as calculated for structures belonging to low energy fragments on Ramachandran map (shown in a) together with the approximation function (2nd degree polynomial). (c) the Ramachandran map with points representing the structures accordant with the approximation function shown in b). (d) the ellipse path assumed to represent the limited conformational sub-space for early-stage intermediate. (e) the ellipse path linking all secondary structures area



the center of the ellipsoid and the concentration reaching zero on the surface of the “drop”, according to the Gauss function:

$$\tilde{H}_t = \frac{1}{\tilde{H}_{tsum}} \exp\left(\frac{-(x_j - \bar{x})^2}{2\sigma_x^2}\right) \exp\left(\frac{-(y_j - \bar{y})^2}{2\sigma_y^2}\right) \exp\left(\frac{-(z_j - \bar{z})^2}{2\sigma_z^2}\right) \quad (2)$$

where $\bar{x}, \bar{y}, \bar{z}$ are the coordinates of the geometric center of the molecule (usually located in the origin of the coordinate system). This is why these values can be considered equal to zero. The size of the molecule is expressed by the triple

$\sigma_x, \sigma_y, \sigma_z$, which is calculated for each molecule individually provided that the orientation of the molecule with the longest possible inter-effective atoms distance is determined according to the appropriate coordinate system axis. The σ values are calculated as 1/3 of the longest distance between two effective atoms calculated along each axis. The value of the Gauss function at any point of protein body is treated as the idealized hydrophobic density defining the hydrophobic core.

The idealized hydrophobicity at any point of the “fuzzy oil drop” can be calculated according to the Gauss function for the molecule located with its geometric center as the

origin of the coordinate system. On the other hand, the

empirical hydrophobicity distribution is calculated according to the function presented by Levitt [19].

$$\tilde{H}o_j = \frac{1}{\tilde{H}o_{sum}} \sum_{i=1}^N (H_i^r + H_j^r) \begin{cases} \left[1 - \frac{1}{2} \left(7 \left(\frac{r_{ij}}{c} \right)^2 - 9 \left(\frac{r_{ij}}{c} \right)^4 + 5 \left(\frac{r_{ij}}{c} \right)^6 - \left(\frac{r_{ij}}{c} \right)^8 \right) \right] & \text{for } r_{ij} \leq c \\ 0 & \text{for } r_{ij} > c \end{cases} \quad (3)$$

where N expresses the number of amino acids in the protein (number of grid points), \tilde{H}_i^r expresses the hydrophobicity of the i th residue according to the accepted hydrophobicity scale (the Aboderin scale was applied in this work [20]), r_{ij} expresses the distance between the i -th and j -th interacting residues, and c expresses the cutoff distance, which according to the original paper [19] is assumed to be 9 Å. The values of $\tilde{H}o_j$ are standardized by dividing them by the coefficient $\tilde{H}o_{sum}$, which is the sum of all hydrophobicities attributed to grid points.

Hydrophobicity distribution in the molecule under consideration appeared to be highly consistent with the idealized one. However, the irregularities observed in many proteins appeared to be target-oriented and related to active sites, such as ligand binding sites or enzymatic active sites.

Kullback-Leibler information entropy

The accordance between the idealized and the observed hydrophobicity distribution is measured according to the Kullback-Leibler relative (divergence) entropy [21], which quantifies the distance between two distributions. The distance between the observed and the theoretical (O/T) distribution was calculated. This value can be estimated only with respect to other solutions. The random distribution of hydrophobicity represented the border case for which the distance (O/R) was calculated. The relation $O/T < O/R$ was taken as evidence for a non-random distribution close to theoretical one.

$$D_{KL}(p|p^0) = \sum_{i=1}^N p_i \log_2(p_i/p_i^0), \quad (4)$$

where: D_{KL} – distance entropy, p – probability of a particular observed event, p^0 – probability in reference distribution. The index “ i ” denotes a particular amino acid. N denotes the number of amino acids in the polypeptide chain.

Results

The structural analysis of the mutants is performed in respect to the ES and LS structural characteristics using the VR model and “fuzzy oil drop” model with the distance entropy applied

to quantitative measurements of the structural differences between two structures under consideration.

Structural analysis of proteins under consideration

A structural analysis of proteins under consideration with respect to the ES and LS is presented in Table 1.

Applicability of the ES model

According to the ES model, structure is generated according to backbone preferences in terms of the V-angle and R-radius of curvature. This is why the values of V-angle and R-radius of curvature (in logarithmic units) as they appear in the crystal structures of proteins under consideration were analyzed versus the idealized curve. The D distance between the projected and observed values of parameters was calculated. It was arbitrarily assumed that proteins with average D below 1 exhibit a structure consistent with the model. However, in view of the availability of the final (LS stage) structures, a $D_{average}$ value above 1 does not imply that the model is inadequate. A low value of D suggests that the structural elements characteristic of the ES structural form have been preserved to a large degree in native (LS structure). All helical fragments are present in both the ES and the LS. That is why low values of $D_{average}$ may suggest a large participation of secondary structures of the helical type.

Two proteins representing extreme cases (large and low D values) are shown as examples in Fig. 2. The distribution of the observed values (V, $\ln(R)$) in comparison to the idealized approximation curve is shown in Fig. 2.

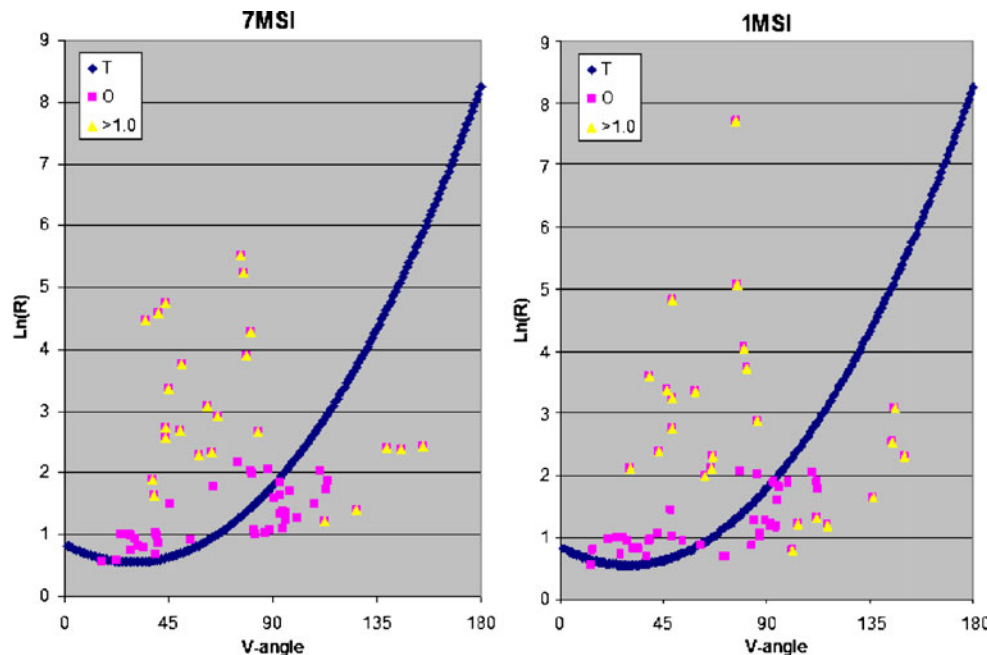
The 3-D structures with residues with $D_{average}$ above 1 are marked in red in this picture in order to visualize the character of the structural motif which is not consistent with the adopted model (Fig. 3).

The accordance of the crystal structure with the ES model is not typically expected. On the other hand, the crystal structure is usually consistent with the LS model, the ES to LS transition is the change of optimal backbone conformation toward the presence of a hydrophobic core. Thus, it is obvious that ES characteristics may be lost in the LS intermediate, although this is not always the case. 1J5B is the only example among the discussed antifreeze proteins

Table 1 The ES and LS characteristics of proteins under consideration. The position of mutations is given in the second column, followed by the value of D (distance), which is a measure of the accordance with the adopted model of ES intermediate. The $D_{average}$ expresses the mean distance between the projected and observed values of parameters that describe the structure of ES intermediate. The proteins with $D_{average}$ values below 1.0 are considered consistent with the ES model. The protein of the relation $O/T < O/R$ is interpreted as accordant with LS model O/T denotes the Kullback-Leibler entropy calculated for the observed (O) distribution of hydrophobicity density and theoretical one (T) treated as the target distribution in comparison with the O/R expressing the distance between observed one and random (R) treated as the target distribution. Chains A were taken for analysis in NMR technique determining the protein structure. The values given in bold denote the case of accordance with appropriate model

PDB - ID	Mutation	D average	O/T	O/L
1AME	P64A, P65A	1.214	0.058	0.066
2AME	P64A, P65A, N14Q	1.174	0.062	0.072
3AME	P64A, P65A, Q9T, Q44T	1.102	0.055	0.060
4AME	P64A, P65A, T18A	1.164	0.061	0.066
6AME	P64A, P65A, M21A	1.202	0.059	0.069
7AME	P64A, P65A, T15A	1.354	0.058	0.067
8AME	P64A, P65A, N14S, A16H	1.251	0.059	0.066
9AME	P64A, P65A, S42G	1.403	0.060	0.068
1MSI	P64A, P65A	1.190	0.065	0.066
2MSI	A16M	1.251	0.080	0.062
3MSI	A16H	1.334	0.077	0.059
4MSI	A16T	1.396	0.068	0.061
5MSI	A16C	1.318	0.071	0.061
6MSI	A16R	1.303	0.067	0.057
7MSI	A16Y	1.294	0.066	0.060
8MSI	P64A, P65A, N14S, Q44T	1.192	0.056	0.062
9MSI	P64A, P65A, T18N	1.238	0.054	0.062
1MSJ	P64A, P65A, T15V	1.237	0.060	0.068
2MSJ	P64A, P65A, N46S	1.163	0.058	0.064
1JAB	P64A, P65A, T18A	1.153	0.058	0.067
1JIA	P64A, P65A, K61I	1.213	0.060	0.064
1B7I	P64A, P65A, K61R	1.275	0.055	0.063
1B7J	P64A, P65A, V20A	1.273	0.057	0.067
1B7K	P64A, P65A, R47H	1.315	0.055	0.066
1KDE	INS(M0) P64A, P65A, INS(K66, D67, E68, L69)	1.113	0.061	0.091
1KDF	INS(M0) P64A, P65A, INS(K66, D67, E68, L69)	0.989	0.059	0.076
2SPG	T15S	1.381	0.060	0.067
1J5B	T(2,13,24,35)V A(7,29)K, A(11,33)E	0.192	0.349	0.072

Fig. 2 The ES model applicability to 7MSI and 1MSI – proteins of lower and higher (respectively) discordance with the assumed model although both of them are treated as representing the structure not accordant with the ES model ($D_{average}$ above 1.). The dark blue symbols – theoretical dependence between V-angle and $\ln(R)$, pink squares – observed parameters and yellow triangles – the residues of the higher than 1.0 unit difference between expected and observed values of $\ln(R)$ for particular V-angle



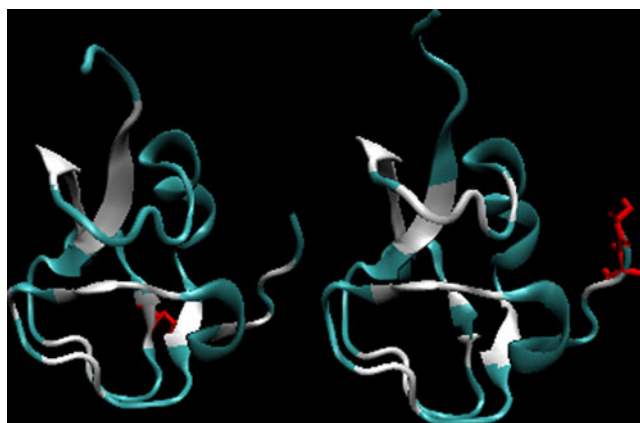


Fig. 3 The 3-D presentation of the 7MSI (left) and 1MSI (right) proteins differing their lower and higher (respectively) accordance with the LS model. The fragments marked in white – residues of difference higher than 1.0 unit shown in Fig. 1 as yellow triangles. The residues shown in red – mutations versus the wild type

(type I). Its structure is entirely helical, and appears to be highly consistent with the ES model. The distribution of hydrophobicity in this molecule is much closer to the random distribution than to the Gaussian one.

Applicability of the LS model

The LS model assumes that hydrophobicity distribution in the protein molecule is consistent with the idealized one, expressed by the three-dimensional Gauss function. The profile showing the hydrophobic interactions collected by effective atoms of each residue as the effect of interactions with other amino acids is shown in Fig. 4.

The 3-D presentation of protein molecules with residues (marked in white) with strongest hydrophobic interactions (responsible for the generation of the hydrophobic core) in two proteins selected to represent the best and the worst accordance with the model under consideration is shown in Fig. 5.

The Kullback-Leibler distance entropy

The accordance between the observed and the idealized hydrophobic density distributions was expressed quantitatively using the Kullback-Leibler distance entropy (as shown in [Materials and methods](#)). The values measuring the distance between the observed and idealized (O/T) and the observed and the random (O/R) distributions are given in Table 1. The analysis of these values suggests that the structural changes do not influence the status of the structure (accordance with the idealized model is preserved). Some proteins undergo changes that result in structure no longer consistent with the adopted model, which suggests that the mutations destroy the hydrophobic core responsible for stabilizing the molecule.

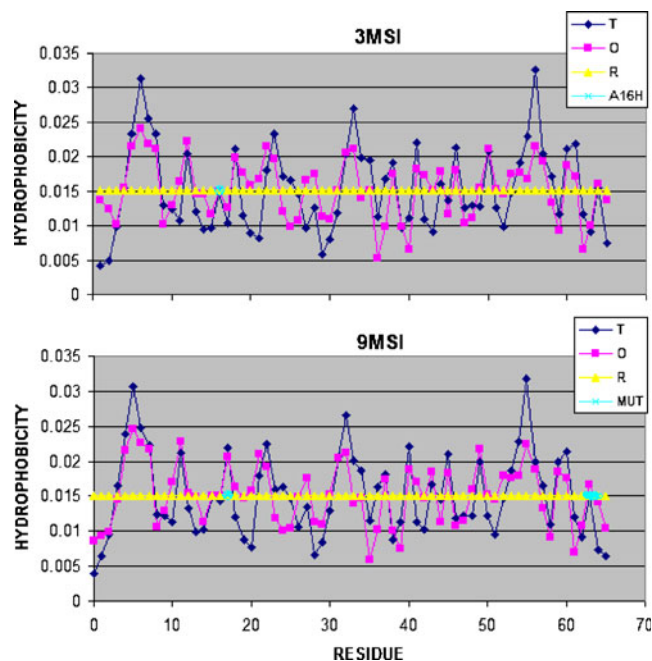


Fig. 4 The hydrophobic density profile for 3MSI and 9MSI showing the idealized and observed distributions. The proteins were selected to show the lowest and the highest respectively accordance between the idealized (T) and observed (O) hydrophobicity distribution. The yellow line shows the random distribution (R). The residues mutated versus the wild type are shown by cyan circles

A particular mutation in position 16 in 2MSI to 7MSI respect to 1MSI appeared to affect the hydrophobic core to such a large extent that it lost its initial structure and became inconsistent with the idealized core structure.

Substituting Pro in positions 64 and 65 with Ala, which is absent in the other investigated proteins and their mutants, suggests that prolines play a critical role as far as hydrophobic core generation is concerned.

The investigated molecules are classified in Table 2 depending on accordance with ES and LS models.

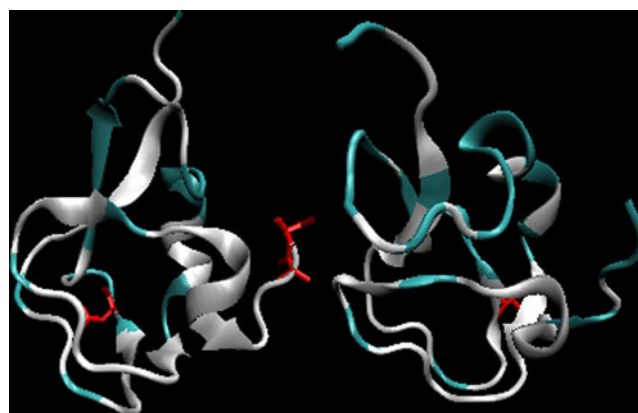


Fig. 5 3-D presentation of 3MSI (left) and 9MSI (right) with the residues of hydrophobicity density differing more than 0.004 versus the expected one given in white. The residues shown in red – the mutated residues

Table 2 Protein classification with respect to the criteria describing/defining the early stage (ES) and late stage (LS) intermediates

		ES model	
		Consistent	Non-consistent
LS model	Consistent	1KDF	1AME, 2AME, 3AME, 4AME, 6AME, 7AME, 8AME, 9AME, 1KDE, 1MSI, 8MSI, 9MSI, 1MSJ, 2MSJ, 2SPG, 1JIA, 1JAB, 1B7I, 1B7J, 1B7K
	Non-consistent	1J5B (I)	2MSI, 3MSI, 4MSI, 5MSI

The majority of the proteins under consideration are very similar (both in terms of sequence and structure), there is only one (1KDF – minimized averaged NMR structure) that satisfies the conditions of both models (ES and LS). This may suggest that the initial ES intermediate was not destroyed in the transition to LS.

The accordance with the LS model is the strongest one in 1KDE structure. The structural fluctuation of dynamic forms seems to be limited by the stabilization imposed by the hydrophobic core (in accordance with the three-dimensional Gauss function).

On the other hand, its four mutants (2MSI, 3MSI, 4MSI, 5MSI) are examples in which mutation prevented the formation of hydrophobic core, which is present in all other structural forms of other mutants of this protein.

Structural differences in pair-wise comparison

A comparison of the intensity of structural changes upon mutation in relation to other proteins of the same group is shown in Table 3. Such a ranking allows contrastive analysis, even more significantly so in this case due to identical (or similar) polypeptide chain length.

The LS model based comparative structural analysis was performed using the Kullback-Leibler divergence entropy treating one of the compared proteins as the target. The

Table 3 Pair-wise comparison of selected mutants (AMI). The values under the diagonal – the RMS-D measurements: the values above the diagonal present the D_{KL} distance entropy between two proteins (according to the column and row headers)

		D_{KL}							
		AMI1	AMI2	AMI3	AMI4	AMI6	AMI7	AMI8	AMI9
RMS-D	AMI1		0.294	0.196	0.201	0.250	0.315	0.294	0.265
	AMI2	0.080		0.267	0.200	0.204	0.150	0.142	0.178
	AMI3	0.078	0.128		0.162	0.208	0.292	0.276	0.227
	AMI4	0.049	0.095	0.074		0.162	0.237	0.213	0.168
	AMI6	0.082	0.046	0.122	0.093		0.240	0.244	0.199
	AMI7	0.056	0.088	0.088	0.058	0.084		0.147	0.226
	AMI8	0.056	0.090	0.096	0.066	0.084	0.051		0.198
	AMI9	0.054	0.099	0.082	0.055	0.098	0.062	0.073	

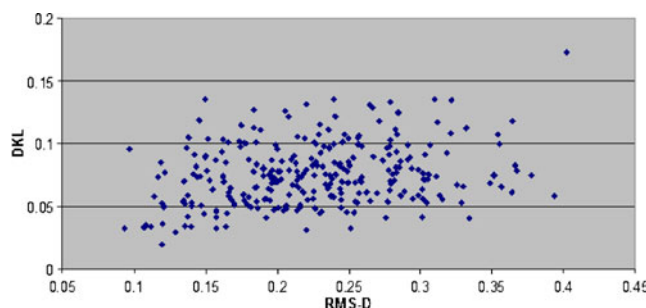


Fig. 6 The relation between traditional similarity measurements expressed as RMS-D values and D_{KL} measurements. The correlation coefficient calculated is equal to 0.2268 with the statistical significance on the level $p < 0.0001$

values received on the basis of these calculations were compared with traditionally used similarity scale expressed by RMS-D values. The appropriate values for selected mutants (group AMI) are given in Table 3.

The correction coefficient for D_{KL} versus RMS-D as calculated using STATISTICA program is equal 0.2268 with $p < 0.0001$. The graphic presentation of this relation is shown in Fig. 6.

Conclusions

The molecules presented in this paper are examples of proteins with structure which seems to satisfy the adopted model of “fuzzy oil drop”. When folding, these molecules satisfy all the conditions defined by non-bonding interactions with simultaneous hydrophobic core formation. Hydrophobic residues located in the central part of the molecule and exposure of hydrophilic residues on the surface are the main tenets of the “oil drop” model introduced by Kauzmann [18]. The Kullback-Leibler entropy [21], which is a measure of the distance between the target distribution (idealized one) and the one observed in a particular molecule revealed good accordance of the observed hydrophobicity distribution with the idealized one.

The Kullback-Leibler entropy calculated for different mutants seems to quantitatively express the scale of structural differences in terms of the hydrophobic core structure.

The selected proteins are examples supporting the reliability of the “fuzzy-oil-drop” model. This model reproduces/imitates the mechanism of protein folding. The modification of the “fuzzy oil drop” model for proteins that are not consistent with this model is under consideration.

The loss of the accordance with the ES model in the LS step of protein folding is obvious, although some proteins with highly preserved secondary structures also exhibit this accordance in their late stage structural form.

It is difficult to verify the applicability of the presented model with respect to biological activity of the proteins under consideration. Their biological function requires high solubility, but no specific interactions understood as necessary formation of binding sites. The antifreeze proteins interact non-specifically and their role is to neutralize water's tendency to be highly organized. The exposure of poorly hydrophobic (i.e., hydrophilic) residues on the protein surface very likely ensures such an effect.

The application of the presented model to the proteins with well-defined active sites may also reveal its ability to locate them. When used for mutants it may estimate the influence of mutation on the potential loss of biological activity [22]. The position of mutation and its relation to the location of residues engaged in biological function may easily be visualized when the $\Delta\tilde{H}$ profile is presented ($\Delta\tilde{H}$ expresses the difference between expected and observed hydrophobicity revealing the residues of significant difference between observation and the model). Such an analysis was presented in [22].

The influence of mutation on the structure and, subsequently, on biological activity was defined using the hydrophobic density distribution.

When hydrophobicity distribution in the protein molecule is consistent with the idealized one, the protein molecule exhibits high solubility, but no specific biological activity. It had been assumed in the past that such proteins with no biological function do not exist. However, the antifreeze proteins appeared to satisfy the above-mentioned conditions. That is why proteins from this group were selected as examples to visualize different forms of the accordance between the assumed model and the real structure in antifreeze proteins.

The pair-wise differences for mutants appeared of much higher magnitude in terms of the relation between the idealized and observed hydrophobicity distributions.

The opposite situation is observed in the group of peroxidases, where the pair-wise comparison reveals far smaller differences.

This paper was focused on good applicability of the Kullback-Leibler entropy as a measure of distance between two distributions.

This method is very simple and it seems to be a suitable tool for automatic analysis of large amounts of data (structures of mutants and/or structures of proteins with equal numbers of amino acids in polypeptide chains).

The protein 3BDN was taken to estimate the applicability to the larger proteins (above 200 amino acids) [23].

The applicability of Kullback-Leibler entropy for the set of proteins belonging to the antifreeze proteins revealed the high accordance of the structure characteristics of this group of proteins with the “fuzzy oil drop” model. It suggests that the hydrophobic core in proteins under consideration represents the structure (hydrophobic density distribution) of three-dimensional Gauss function. The consequence of this observation is that the presence of external force field in folding process simulation may be treated as the heuristic model for protein folding simulation. The other group of proteins were also recognized as proteins of structure in accordance with “fuzzy oil drop” model. They are: fast folding proteins, cold shock proteins and some proteins in the form of homodimers (currently under consideration). The protein of the structure assumed to represent the early stage step of folding process and its native structural form appeared to be well accordant with both ES and LS mode respectively [24]. The “fuzzy oil drop” model is able to explain the structural differentiation of two homologous proteins of significantly different structure (change of α -helix to the β -structural form). Although all proteins listed as accordant with the “fuzzy oil drop” model are of the category “easy predictable” (according to CASP classification [25]) the meaning of the presented model is its general character. The introduction of external force field and the accordance of structures of some proteins with the model suggests the significant role of the environment for folding process.

Acknowledgments This work was financially supported by Jagiellonian University – Medical College - Grant No K/ZDS/001531.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Hari SB, Byeon C, Lavinder JJ, Magliery T (2010) *Protein Sci* 19:670–679
2. Yagawa K, Yamano K, Oguro T, Maeda M, Sato T, Momose T, Kawano S, Endo T (2010) *Protein Sci* 19:693–702
3. Chung HS, Shandiz A, Sosnick TR, Tokmakoff A (2008) *Biochemistry* 47:13870–13877
4. Sapra KT, Balasubramanian GP, Labudde D, Bowie JU, Muller DJ (2007) *J Mol Biol* 376:1076–1090

5. Couñago R, Wilson CJ, Peña MI, Wittung-Stafshede P, Shamoo Y (2008) *Protein Eng Des Sel* 21:19–27
6. Kurpiewska K, Font J, Ribó M, Vilanova M, Lewiński K (2009) *Proteins* 77:658–669
7. Lonquety M, Lacroix Z, Papandreou N, Chomilier J, Lonquety M, Lacroix Z, Papandreou Z, Chomilier J (2004) *Nucleic Acids Res* 37:D374–D379
8. Fernández M, Caballero J, Fernández L, Abreu JI, Acosta G (2008) *Proteins* 70:167–175
9. Capriotti E, Fariselli P, Casadio R (2005) *Nucleic Acids Res* 33: W306–W310
10. Karchin R, Kelly L, Sali A (2005) *Pac Symp Biocomput* 2005:397–408
11. Fernández L, Caballero J, Abreu JI, Fernández M (2007) *Proteins* 67:834–852
12. Hoekstra HE, Coyne JA (2007) *Evolution* 61:995–1016
13. Christ D, Winter G (2003) *Proc Natl Acad Sci USA* 100:13202–13206
14. Roterman I, Brylinski M, Konieczny L, Jurkowski W (2007) In: de Brevem AG (ed) *Recent advances in structural biology. Research Signpost, Trivandrum, India.* ISBN 978-81-308-0208-4
15. Roterman I (2009) *Structure-function relation in proteins. Transworld Research Network T.C. 37/661(2), Fort P.O. Trivandrum, Kerala, India*
16. Roterman I (1995) *J Theor Biol* 77:283–288
17. Konieczny L, Brylinski M, Roterman I (2006) *In Silico Biol* 6:15–22
18. Kauzmann W (1959) *Adv Protein Chem* 14:1–63
19. Levitt M (1976) *J Mol Biol* 104:59–107
20. Aboderin A (1971) *Int J Biochem* 2:537–544
21. Nalewajski RF (2006) *Information theory of molecular systems. Elsevier, Amsterdam.* ISBN 978-0-444-51966-5
22. Prymula K, Sałapa K, Roterman I (2010) *J Mol Model* 16:1269–1282
23. Stayrook SE, Jaru-Ampornpan P, Ni J, Hochschild A, Lewis M (2008) *Nature* 452:1022–1025
24. Religa TL, Markson JS, Mayor U, Freund SM, Fersht AR (2005) *Nature* 37:1053–1056
25. Orengo CA, Bray JE, Hubbard T, LoConte L, Sillitoe I (1999) *Proteins Suppl* 3:149–170

Theoretical study of the surface properties of poly(dimethylsiloxane) and poly(tetrafluoroethylene)

Andrea Michalkova · Sonia Tulyani · James Beals · Jerzy Leszczynski

Received: 1 February 2011 / Accepted: 22 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract Molecular dynamics (MD) simulations of poly(dimethylsiloxane) (PDMS) and poly(tetrafluoroethylene) (PTFE) were carried out to determine their surface properties and energies. This study helps to gain better insight into the molecular modeling of PDMS and PTFE, in particular how different approaches affect calculations of surface energy. Current experimental and theoretical data were used to further understand the surface properties of PDMS and PTFE as well as to validate and verify results obtained from the combination of density functional theory (DFT) calculations (including periodic boundary conditions) and MD simulations. Detailed analysis of the structure and electronic properties (by calculation of the projected density of states) of the bulk and surface models of PDMS and PTFE was performed. The sensitivity of the surface energy calculation of these two polymers to the chemistry and model preparation was indicated. The balance between the molecular density, weight (which also reflects bond orientation in the surface region), bond flexibility, and intramolecular interactions including bond stretching was revealed to govern the results obtained. In modeling, the structural organization of polymer near a given surface (types and number of end groups and broken bonds due to application of different cut offs of the periodic structure) also significantly affects the final results. Besides

the structural differences, certain simulation parameters, such the DFT functionals and simulation boxes utilized, play an important role in determining surface energy. The models used here were shown to be sufficient due to their good agreement with experimental and other theoretical data related to surface properties and surface energies.

Keywords Poly(tetrafluoroethylene) · Poly(dimethylsiloxane) · Vienna ab initio simulation package · Surface energy

Introduction

The widespread use of aluminum for a variety of applications has resulted in the development of a large number of surface treatments for this metal. Chemical pretreatment regimens can be used for example to vary the degree of adhesive bonding [1]. Reducing the adhesion between ice and metallic surfaces is important for a broad range of applications including aircraft and heat exchanger performance. To protect the aluminum (alumina) components from the effects of water (ice), a layer of coating material is applied at the surface. Numerous commercially available materials, coatings, and paints are advertised to have low friction or non-stick properties. Among them are poly(dimethylsiloxane) (PDMS) and poly(tetrafluoroethylene) (PTFE) [2].

PDMS is a macromolecule with a $[-O-Si(CH_3)_2-]_n$ chemical formula where n is the number of repeat units or monomers [3]. Each carbon and its bonded hydrogen atoms are treated as a single united particle, whereas Si and O atoms are represented as individual particles. An extensive study of the structural and dynamic properties of the PDMS united atom model can be found in ref [4]. PDMS possesses

A. Michalkova · J. Leszczynski (✉)
Interdisciplinary Nanotoxicity Center (INC), Department of
Chemistry and Biochemistry, Jackson State University,
1400 Lynch Street, P.O. Box 17910, Jackson, MS 39217, USA
e-mail: jerzy@icnanotox.org

S. Tulyani · J. Beals
United Technologies Research Center (UTRC),
411 Silver Lane,
East Hartford, CT 06108, USA

an extremely low viscosity caused by its chain flexibility, weak intermolecular interactions, and the relatively free motion of its side groups [5]. PDMS is one of the important industrial silicon polymers with many different applications in, for example, adhesives, biomedical devices, coatings, lubricating oils, contact lenses, and heat resistant tiles [6–8].

PTFE (also known as Teflon) is a homopolymer of tetrafluoroethylene ($\text{CF}_2=\text{CF}_2$). Numerous experimental [9–11] and simulation [12] studies have been performed on crystalline PTFE because of its extensive industrial use [13]. Due to its attractive surface/interface properties such as water and oil repellency (extreme hydrophobicity), solvent resistance, low surface tension, thermal stability, low friction coefficient, high lubricity and low dielectric constant, PTFE has many applications as a coating material, anesthetic, surfactant, oxygen carrier and sealant.

Surface properties determine material properties such as mechanical strength, friction, and shape. The properties of the surface govern interactions with a second material, adhesion properties, and the hydrophilicity/hydrophobicity of the material. Therefore, gaining a molecular-level understanding of the above mentioned substrates is important in explaining their unique physical properties, specifically their surface energy. Furthermore, such information can help in understanding the interfacial interactions with water (ice) in development of the associated physically based model(s) used to predict the mechanical strength of the adhesive bond between ice and substrates of interest.

The surface properties of simple polymers have been investigated in many molecular dynamics (MD) studies [12, 14–17]. However, the investigation of PTFE has been limited due to the absence of good force fields. Several works have been published on the improvement of force fields based on semiempirical methods or quantum chemistry calculations [18–23]. Among these studies, the force field of Jang et al. [18] was found to successfully describe the conformational characteristics of model perfluoroalkanes obtained from experiments and theoretical calculations. Moreover, they were used and well recommended in a study on the characteristics of PTFE melts [24]. Several theoretical works have been published regarding the structural features of PDMS [25–30] and PTFE [22, 31, 32]. All these latter studies performed force field calculations and present the results of investigation of bond lengths, bends and torsions.

There are also several publications on the surface properties [33–35] and surface energy values of PDMS [36, 37, for a list of all references see 38], and the surface properties [24, 39, 40] and surface energy values of PTFE [for a list of references see 41–44]. In most of these studies, the surface energy values were obtained experimentally by measurement of the contact angle, except in two studies in which spectroscopic ellipsometry [37] or a

new ink stamp technique using a soft mold [36] were applied. Theoretical studies of structure and surface properties using MD were applied to investigate the interfacial structure and dynamics of PDMS-vapor and PDMS-water systems [29], PDMS melts near silica surfaces [30], and the surface characteristics for PTFE melts [24]. Moreover, MD was applied to calculate the interface between amorphous poly(methylmethacrylate) (PMMA) and PTFE [39], to simulate the transport properties of water-ethanol mixture [28] and polyhedral oligomeric silsesquioxanes in PDMS [27], and to examine the behavior of polymers under high pressure [35]. A classical force field for PDMS and its oligomers [45], for poly(vinylidene fluoride) (PVDF) and its oligomers [31], and for perfluoroalkanes and PTFE [22] has been derived on the basis of intermolecular binding energies, molecular geometries, molecular electrostatic potentials, and conformational energies obtained from quantum chemistry calculations on these model compounds. Transferability of these force fields to longer perfluoroalkanes was concluded [22]. An all-atom force field for computer simulation of polysilanes, including their alkyl and phenyl side-chain-state simulations of simple silane molecules, was developed [25]. A commercially available force field was modified to represent the equilibrium solid state of PTFE and used to examine various aspects of disorder in its phase transformation [32].

The present study constructed a model of the selected materials and studied the electronic and surface characteristics that affect the surface energy and mechanical properties. Experimental validation of the model predictions will be discussed. MD calculations were used to establish the surface energies, which was the main goal of this study. The work presented here is also associated with the molecular modeling effort to predict the mechanical response of studied materials. The first step of this effort will focus on the comparison of theoretical and experimental structural characteristics. For this reason, we have selected two substrates with widely varying structures. The second step will be to study the electronic properties and vibrational infrared (IR) spectra as well as to investigate the development of the empirical relationship between the structures of the substrate and the calculated surface energies.

Simulation methods

The calculations were performed using the Vienna *ab initio* simulation package (VASP) [46–48] to obtain the relaxed structures and corresponding electronic energies for all studied systems. VASP is based on fully self-consistent density functional theory (DFT) calculations to solve the

Kohn–Sham equations [49] with implementation of projector-augmented wave (PAW) atomic pseudopotentials [50, 51] and a plane wave basis set. The following procedure was performed to obtain the equilibrium structures of the studied substrates. The localized density approximation (LDA) parameterized according to Perdew–Zunger [52], and the generalized gradient approximation (GGA) parameterized according to Perdew–Wang [53] were employed for the description of the electron exchange–correlation interaction. The PAW method improves transferability and reduces the number of plane-waves required in the expansion of the Kohn–Sham orbitals in order to improve computational efficiency.

This is the first study on surface energy calculations of PTFE and PDMS. The LDA and GGA approaches have been applied widely as standard methods to study the surface properties of many different substrates including polymers [54–56], and have been shown to correctly represent experimental trends in their structure and properties [57–59], successfully predict the spectrum of PTFE [60], and to yield good relative bond energies for highly coordinated atoms and surface energy for amorphous silicon [61].

All simulations were done at room temperature (298.15 K). In the dynamics calculations a canonical ensemble with a Nose’ thermostat procedure [62] was used. The Verlet velocity algorithm [63] with a time step of 1 fs was chosen, and a simulation time of 2 ps was applied.

A plane-wave cutoff energy of 400 eV (implying a high-precision quality of the calculation) was applied in all calculations. K-space sampling was employed using the scheme of Monkhorst and Pack [64]. The $[1 \times 1]$ unit cell of the polymer structure was used to prepare surface and bulk PDMS and PTFE models. An initially Γ -centered Monkhorst–Pack $1 \times 1 \times 1$ grid was applied for the Brillouin zone sampling for bulk and surface PDMS and PTFE structures. Moreover, different K-points were applied in order to study their influence on energetics. The original $1 \times 1 \times 1$ K-point mesh was substituted by $2 \times 2 \times 2$ and $3 \times 3 \times 3$ K-point mesh used for the bulk models, and $2 \times 2 \times 1$ and $3 \times 3 \times 1$ grids were employed for the surface models of PDMS and PTFE (the last K-point number was decreased proportionally corresponding to the reduction of the Brillouin zone). The MD simulations were carried out first for the above mentioned boxes of PTFE and PDMS using a $1 \times 1 \times 1$ grid and then using all (above mentioned) larger grids.

The convergence criterion in the electronic cycle (i.e., computation of the electron probability density) was set to 10^{-5} for the bulk and surface slabs. Using the same pseudopotentials, energy cut-offs and other simulation parameters are essential for correct comparison of the studied materials.

Substrate models

Amorphous PDMS and PTFE polymer structures were simulated using a 3D cubic simulation box with periodic boundary conditions in all directions. First, the equilibrium bulk phase was modeled. The bulk models of PDMS and PTFE were mimicked, applying the supercell approach for the main supercell.

The models were built by applying the Amorphous Builder software package [65]. The models contain different amounts of molecules per the $[1 \times 1]$ unit cell that was used for all models. This modeling is based on the density of the polymer per each bulk according to the following algorithm: using a simple polymer chain, the software selects a fragment of the substrate and creates a simulation box by cutting out the polymer structure based on the appropriate bulk density and number of monomers in the unit cell. The possibility of a wide selection of several different ways how the simulations boxes can be cut, leads to several potential various unrelaxed models of the bulk and surface of PTFE and PDMS.

The initial PDMS and PTFE configurations were generated so that the total number of PDMS and PTFE molecules achieves the experimental bulk density (the following experimental densities were used for both polymers: 0.965 g cm^{-3} for PDMS [66, 67] and 2.2 g cm^{-3} [68, 69]). The PDMS simulation box consists of four monomers with total molecular weight of 682 g mol^{-1} and unit cell dimensions of $10.5 \times 10.5 \times 10.5 \text{ \AA}$. The initial unit cell of PTFE contains eight monomers with a total molecular weight of 888 g mol^{-1} and cell dimensions of $8.75 \times 8.75 \times 8.75 \text{ \AA}$. The periodic boundary condition was imposed in all three (x-, y- and z-) directions. Figure 1 illustrates the structure of bulk model of PTFE and PDMS with $[1 \times 1]$ simulation box as modeled using the Amorphous Builder program [65].

In the next step, a surface cell was constructed so that the bulk phase simulation box of PTFE and PDMS was elongated in the z-direction, which generates free surfaces at both sides. A vacuum spacing of 10 \AA was inserted in order to ensure that the free surfaces (periodic images) would not interact with each other at such a distance. The main reason for applying a 10-\AA elongation is that this vacuum spacing was shown to have a negligible effect on the estimated surface energy of different materials ($<10 \text{ mJ m}^{-2}$ or less than 1%) [70].

To test how different fragment (configuration) of the polymers affects the convergence of the geometry and total energy, different cut offs of the periodic structure were applied to modify the surface and bulk models of PDMS and PTFE. These models possess $[1 \times 1]$ unit cells with the same dimensions and density but have different atoms close to the borders of the simulation box. Therefore, based on

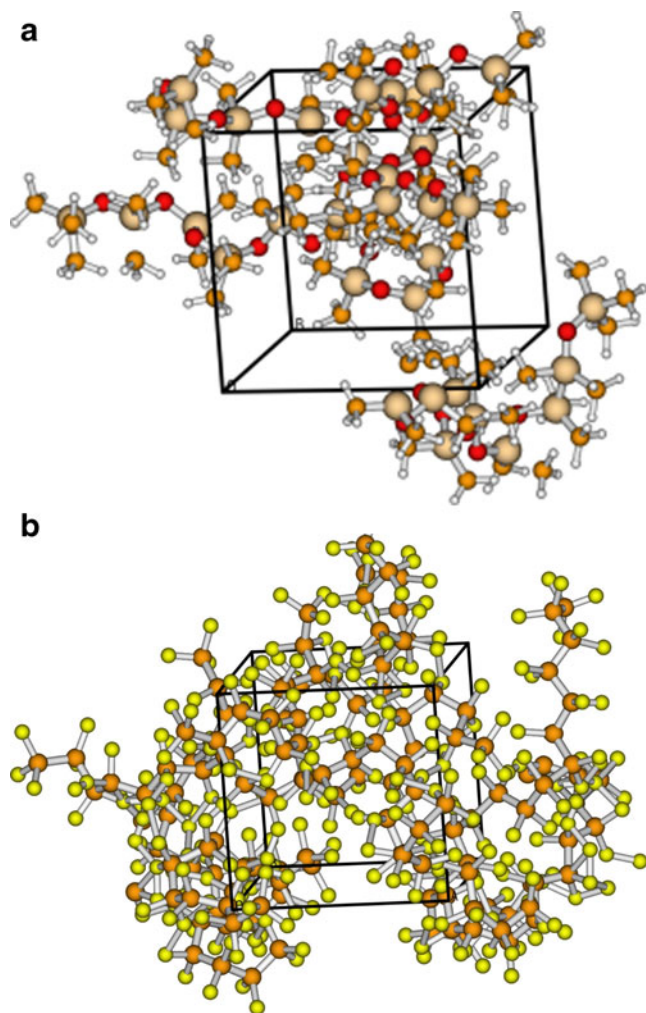


Fig. 1 Structure of the bulk model of poly(dimethylsiloxane) (PDMS) (a) and poly(tetrafluoroethylene) (PTFE) (b) with [1×1] simulation box as modeled using the Amorphous builder program [65]

such differences, one is able to determine how different functional groups and their density close to the surface contribute to changes in the energetics.

The electronic structure and IR vibrational spectra of the isolated studied polymers were also investigated. The projected density of states (PDOS) was calculated for the bulk models of PDMS and PTFE before and after relaxation to see how relaxation affects polarization of the substrates. IR spectra were calculated for the surface and bulk models of PDMS and PTFE and compared with previously published data.

Surface energy calculation

After determining the total energies of the bulk and surface models of PDMS and PTFE by applying the conditions described above, the surface formation energy (E_{se} , the energy required to create the surface) was calculated. E_{se} is typically represented per unit surface area and can be

obtained from the area per surface atom and the number of surface atoms in the unit cell with surface atoms exposed to vacuum on both sides of the slab. It is calculated as the change in energy from bulk to surface divided by the area of each surface [the difference between the total energy of a surface slab $E(\text{surf})$ and the equivalent number of metal atoms (N) in the bulk environment ($E(\text{bulk})$)] using the following equation:

$$E_{se} = E(\text{surf}) - (N \cdot (E(\text{bulk})/n)) / (2 \cdot A) \quad (1)$$

where $E(\text{surf})$ and $E(\text{bulk})$ are the total energies of the slab and the bulk, respectively; N is the number of atoms in the surface cell (simulation box); n is the number of atoms in the bulk cell; and A is the area of the surface. The factor of 2 in the denominator accounts for the fact that two surfaces are created. In this formulation, the surface energy is described to originate solely from the reduced interaction of the surface atoms with their nearest neighbors.

Only a few theoretical works on the structural properties PDMS and PTFE have been published, and these have been devoted mostly to calculations of force field parameters (see [Introduction](#)). None of these publications calculated the exact $E(\text{surf})$ and $E(\text{bulk})$ values for PTFE and PDMS required to obtain E_{se} directly from Eq. 1. Therefore, in this paper, the total energies of the surface and bulk models of PDMS and PTFE after their complete structural relaxation are calculated and used to estimate the surface energies.

Results

Relaxation of models

First, relaxation of the surface and bulk models of PDMS and PTFE was performed, and the geometrical changes due to the relaxation were analyzed. The structures of bulk PDMS and PTFE are illustrated in Fig. 1. The physical properties of surfaces of different materials depend strongly on their microscopic features [71]. Therefore, the structures of PDMS and PTFE were analyzed at the atomistic level. Tables 1 and 2 list the geometrical parameters (bond lengths, bond angles and dihedral angles) of the bulk and surface models of PDMS and PTFE before and after relaxation.

The bulk models of PDMS and PTFE consist of four $\text{SiO}_2(\text{CH}_3)_2$ monomers and eight C_2F_4 monomers, respectively. The dominant molecular energy contributions to the formation of free surfaces of PTFE and PDMS are from the van der Waals and Coulombic energy terms. Comparison of the unrelaxed and relaxed structures of the bulk models of PDMS and PTFE indicates that relaxations occur in all directions towards the surface. Most attention is paid to

Table 1 Bond lengths (Å), angles and dihedral angles (°) for unrelaxed and relaxed surface and bulk models of poly(dimethylsiloxane) (PDMS)

Substrate Model/ feature	Unrelaxed		Relaxed		Literature [25, 28, 30]
	Surface	Bulk	Surface	Bulk	
D(Si-C)	1.589–1.740	1.628–1.725	1.717–1.888	1.857–1.882	1.863, 1.899; 1.9073; 1.9
D(Si-O)	1.531–1.613	1.528–1.597	1.545–1.667	1.639–1.667	1.656; 1.64
(O-Si-O)	105–121	106–121	107–123	104–110	110.7; 107.8
(C-Si-C)	74–125	106–126	110–113	107–118	113.19; 114.9; 109.2
(C-Si-O)	92–114	104–134	108–123	105–115	114.9; 110.7
(C-Si-O-Si)	32–100	–21–100	8–73	–12–121	
(O-Si-O-Si)	13–63	55–95	9–81	41–121	

relaxation of the surface-bridging Si–O and S–CH₃ groups of PDMS, and the C–F groups of PTFE, which are expected to make a significant contribution to the total energy and also to play a main role in interactions with adsorbates. The surface atoms are more fluorine than carbon atoms in the case of PTFE and predominantly methyl groups for PDMS. The ratios correspond with the chemical formulae of these polymers, revealing 33% and 67% carbon and fluorine atoms, respectively, in PTFE; and 9%, 8%, and 21% silicon, oxygen and carbon atoms, respectively, in PDMS.

Relaxation leads to much larger structural changes in the bulk models compared with the surface models. Due to steric considerations, the bulk structure prefers to remain more compact than the surface structure (it spreads more in all directions). The difference is accompanied by larger changes in geometrical characteristics. This result confirms the previously mentioned statement that the bulk model plays a key role in the calculation of the surface energy (differences in E_{se} values are more likely due to the relaxation of the bulk than the surface box). Therefore, we suggest that the bulk model structure more strongly affects the adhesive and cohesive properties of the studied coatings.

To help predict the dependency of surface properties on microstructure, the geometrical characteristics of the bulk and surface models of PDMS and PTFE before and after relaxation were analyzed in detail (see Tables 1, 2). The

values obtained were compared with structural data published in other theoretical studies of PDMS [25–30] and PTFE [22, 31, 32] obtained mostly from force field calculations. The bond lengths in these theoretical studies differ by about 0.02 Å (C–F and Si–O), and 0.04 Å (C–C and Si–C), and angles vary in the interval 2–9°. As one can see, all of our results regarding geometrical features agree well with previously published data for both PDMS and PTFE, validating our surface and bulk models of PDMS and PTFE.

The initial and final structures for both PTFE and PDMS differ, while the structures of the surface and bulk models are almost the same. Evaluation of these parameters shows that bond lengths and torsions change significantly in both studied coatings compared with bends. This means that the initial structure was not close to the global minima, and relaxation of these models is important to obtain the correct properties of the studied polymers. The largest increase in distance due to structural relaxation was found for the Si–O bond length of PDMS (increased about 0.13–0.16 Å) and for the C–C bond length of PTFE (increased about 0.08 Å). The largest change in dihedral angles was seen for C–C–C–C in PTFE and for O–Si–O–Si in PDMS simulation boxes (about 30°). The angles vary only slightly. This means that a tetrahedral geometry for the Si–O–C and F–C–F is preferred, and the geometry remains in the sp³ mode. Such relaxed configurations were shown to be the most favorable for both

Table 2 Bond lengths (Å), angles and dihedral angles (°) for unrelaxed and relaxed surface and bulk models of poly(tetrafluoroethylene) (PTFE)

Substrate Model/ feature	Unrelaxed		Relaxed		Literature [22, 30, 31]
	Surface	Bulk	Surface	Bulk	
D(C-C)	1.545–1.585	1.489–1.589	1.520–1.612	1.492–1.586	1.566, 1.573 1.534; 1.53
D(C-F)	1.347–1.385	1.346–1.365	1.283–1.365	1.340–1.374	1.339, 1.351 1.357; 1.36
(F-C-F)	108–111	108–110	110–118	108–112	108.5, 110.1 105.3 109.5
(C-C-C)	115–122	115–121	107–114	115–124	115.6 118.2 109.5
(C-C-F)	106–109	106–111	107–113	106–111	109.5 107.7 109.5
(C-C-C-C)	–97 to –168	–86 to –174	–83 to –166	–55 to –161	
(F-C-C-F)	–145 to –166	–156 to –176	–145 to –174	–153 to –160	

models of both substrates, which is consistent with the results of other theoretical works [22, 25–32].

Comparison of the relaxed structures of different bulk supercells reveals the influence of the initial set up and structure on the final geometry. The intervals of the geometrical parameters that characterize the new models are almost the same as those for the original ones. On the other hand, in the new models all bulk and surface atoms exhibit small shifts in their positions and orientations towards the rest of the atoms in the system. The main difference is caused by the fact that, even after relaxation, different atoms remain close to the borders of the unit cell, which leads to a modified configuration inside the box. The relaxation of bulk of PDMS and PTFE using $2 \times 2 \times 2$ and $3 \times 3 \times 3$ K-point mesh conditions lead to very similar atomic distortions as obtained with $1 \times 1 \times 1$ -type models. Moreover, relaxation and application of the different K-points method leads to the same dominance of atoms in the models as mentioned above. The same is true for the surface as for the bulk simulation boxes.

Projected density of states

The results suggest that the calculated properties (including mechanical) for these coatings relate to their electronic structures, and their different properties originate from a particular binding between the silicon and oxygen in PDMS, and carbon and fluorine in PTFE. In order to better understand these properties, we analyzed electron densities based on the calculations of the PDOS, which are shown for unrelaxed and relaxed PDMS and PTFE bulk models in Figs. 2 and 3. The PDOS results reveal that relaxation leads to a perturbation of the surface—relaxation induces changes in the density profile. The value of the Fermi energy for unrelaxed PDMS and PTFE is -19.63 eV and -22.83 eV, respectively. In the case of relaxed surfaces these values are shifted to -18.85 eV for PDMS and to -24.67 eV for

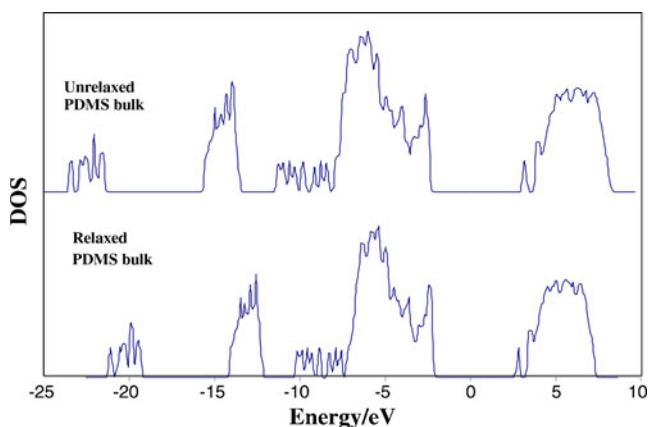


Fig. 2 Projected density of states for the unrelaxed and relaxed bulk models of PDMS

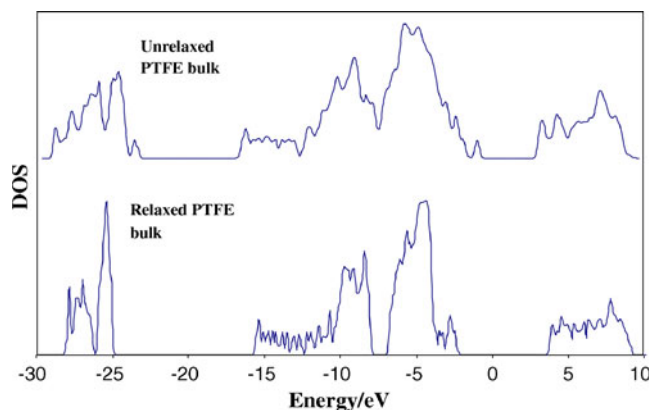


Fig. 3 Projected density of states for the unrelaxed and relaxed bulk models of PTFE

PTFE. This corresponds to a slight shift (about 1 eV) of both bands, conduction and valence, after relaxation compared with the unrelaxed polymers. Both bands are placed in a less negative area for relaxed bulk PDMS model. On the other hand, in the case of the relaxed PTFE bulk model, the valence band is located in a more negative area but the conduction band has moved to smaller negative energy values. The values of the band gaps are 4.58 eV for relaxed bulk PDMS and 8.966 eV for relaxed bulk PTFE, 6.174 eV for unrelaxed PTFE and 3.488 eV for unrelaxed PDMS.

A small contraction of the valence band for both PDMS and PTFE bulk models was found (from 6.6 eV for PTFE before relaxation to 3.53 eV after relaxation, and from 2.97 eV to 2.5 eV for the bulk PDMS model). This change is much smaller in the case of PDMS due to the small differences in the structure caused by the relaxation. If one compares the initial and final optimized structures of bulk PTFE, the difference is more significant, which leads to the shortening of the valence band. For both polymers, relaxation causes large changes in the DOS peaks, which can be due to the appearance of new bonds. The analysis of the conduction band in the total DOS profile reveals the same tendency of changes as described above for the valence band. The width is shortened from 16.79 eV (unrelaxed) to 12.585 eV (relaxed) for PDMS, and from 17.085 eV (unrelaxed) to 13.856 eV (relaxed) for PTFE. Comparing all the changes in valence and conduction bands, the following order is revealed: PTFE-valence (46 %) > PDMS-conduction (25 %) > PTFE-conduction (19 %) > PDMS-valence (16 %). If one considers distinct peaks in the valence band, after relaxation they become less spread (especially in the case of PTFE), which indicates that the chemical bonds are less uniform, especially in the border area. The DOS profiles confirm the changes in the geometrical parameters discussed in the previous section, which relate to differences on the surface and in binding of the atoms in the bulk model. One can conclude that the

relaxation is more significant in the case of PTFE than in PDMS, which largely affects the chemical activity of the surface region of the surface and bulk models. The different behavior of PDMS and PTFE is caused by the different chemistry of these two coatings, including the different nature of intramolecular binding and the different atoms exposed by the two surfaces. PDOS analysis shows that relaxation affects the electronic distribution of the border area of both models, which leads to an increase in total reactivity of substrate. Analysis of DOS of different materials such as Er-doped silicon [69] has also shown that symmetry determines the shape and energy splitting of DOS peaks. This is in agreement with our finding that increasing the symmetry (in our case relaxed structures) causes the peaks to be sharper and more split.

IR spectra

Vibrational spectroscopy is one of the most important methods used to characterize polymers, because it can give information on chemical structures as well as molecular orientation. The vibrational spectroscopy of polysilanes have been published by several authors [72–74] as well as the IR spectra for PTFE [75–78]. The IR spectra of surface and bulk models of PDMS and PTFE were calculated and are presented in Figs. 4 and 5. The spectra agree well with the chemistry of the studied polymers described in the previous two sections. They will be compared with the experimental IR spectra for PDMS and PTFE [79].

The calculated region of the modes of the functional groups $-\text{Si}-\text{CH}_3$, $-\text{Si}-\text{O}-\text{Si}-$, $-\text{CH}=\text{CH}_2(\text{C}=\text{C})$ and $-\text{CH}=\text{CH}_2(\text{C}-\text{H})$ for relaxed bulk of PDMS are shown. The spectrum of the relaxed surface model of PDMS is presented for reference. Experimentally measured vibrational frequencies for the PDMS base monomer are as

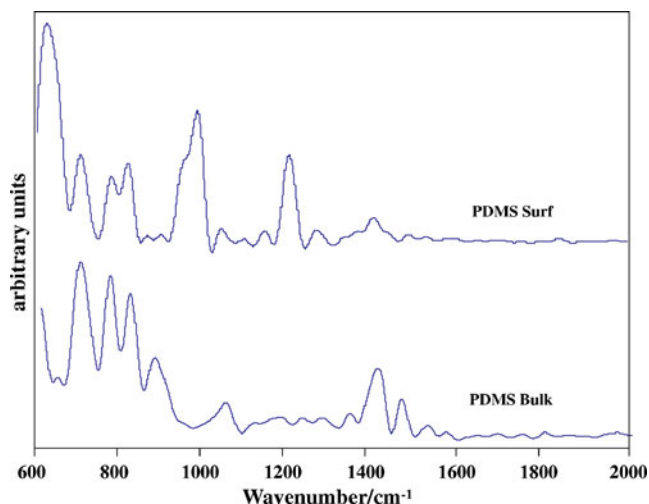


Fig. 4 Infrared spectra of the PDMS surface (Surf) and bulk models after relaxation

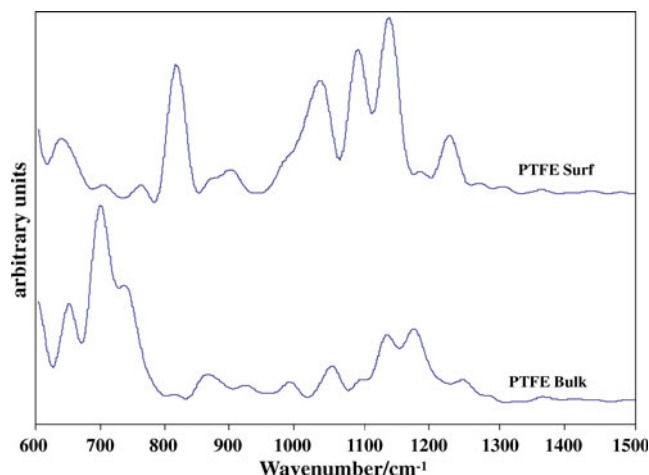


Fig. 5 Infrared spectra of the PTFE surface (Surf) and bulk models after relaxation

follows: $-\text{Si}-\text{CH}_3$ is $1,260$ and 800 cm^{-1} , $-\text{Si}-\text{O}-\text{Si}-$ is $1,011\text{ cm}^{-1}$, $-\text{CH}=\text{CH}_2(\text{C}=\text{C})$ is $1,060\text{ cm}^{-1}$ and $-\text{CH}=\text{CH}_2(\text{C}-\text{H})$ is 960 cm^{-1} [72]. An absorption peak at $1,065\text{ cm}^{-1}$ was also observed for isolated PDMS, which is due to the stretching vibration of the $\text{Si}-\text{O}-\text{Si}$ group [73]. An IR study of vacuum-deposited thin films of PDMS indicates a 760 cm^{-1} mode, which belongs to rocking of the methyl group, and a band around $2,090\text{ cm}^{-1}$ is due to $\text{Si}-\text{H}$ stretching [74]. Moreover, bands in the $1,000\text{--}1,100\text{ cm}^{-1}$ region of $\text{Si}-\text{O}$ bonds are also present, and the vibration in the $900\text{--}700\text{ cm}^{-1}$ region can be assigned to $\text{Si}-\text{C}$ bonds [74]. In our computational study, an $\text{Si}-\text{O}$ mode was found at $1,044\text{ cm}^{-1}$, and a low frequency $\text{Si}-\text{C}$ bond corresponds to 813 cm^{-1} and 945 cm^{-1} for $\text{C}-\text{H}$ bond in the bulk model of PDMS. As one can see, these calculated modes are shifted slightly (to about $20\text{--}30$ and 10 cm^{-1} higher frequencies for $\text{Si}-\text{O}$ and $\text{Si}-\text{C}$, and to about 15 cm^{-1} lower frequency for $\text{C}-\text{H}$) compared with the experimental values of the PDMS base monomer [72]. The calculated interval of $700\text{--}900\text{ cm}^{-1}$ is in good agreement with an experimentally obtained spectrum for PDMS deposited on a KBr window at room temperature [74]. The spectrum of the PDMS surface model contains the same peaks observed for the bulk PDMS model. It differs from the bulk PDMS model in the intensity of the most important modes, which are shifted only slightly ($1,047\text{ cm}^{-1}$ for $\text{Si}-\text{O}$, 817 cm^{-1} for $\text{Si}-\text{C}$ and 983 cm^{-1} for $\text{C}-\text{H}$). However, in the case of the surface model, the $\text{Si}-\text{O}$ peak (which occurs experimentally at $1,011\text{ cm}^{-1}$) and the bulk PDMS model, the $\text{C}-\text{H}$ peak, the experimental frequency of which occurs at 960 cm^{-1} , may be masked by the broad band of $\text{Si}-\text{O}-\text{Si}$.

The mode for rocking of the methyl group is observed in both calculated spectra for the bulk (766 cm^{-1}) and surface (779 cm^{-1}) PDMS models and agrees very well with the observed value [74]. Bands for $\text{Si}-\text{CH}_3$ ($1,268\text{ cm}^{-1}$ for bulk, and $1,272\text{ cm}^{-1}$ for the surface PDMS model) are

found in both spectra, which is consistent with the experimentally observed $1,260\text{ cm}^{-1}$ frequency [72]. Bands close to $1,400\text{ cm}^{-1}$ [79] (not mentioned in previously published experimental studies) are indicated in the spectrum of PDMS for the C–H bonds and correspond well with our findings ($1,393\text{ cm}^{-1}$ for surface and $1,407\text{ cm}^{-1}$ for the bulk PDMS model). Generally, only small shifts between calculated and experimental bands were revealed for both models. Thus, it can be concluded that the models of PDMS used are sufficient to study the vibrational spectra of this polymer.

Figure 5 shows the calculated IR spectra of the bulk and surface models of PTFE obtained at room temperature. Both spectra were practically featureless above $1,400\text{ cm}^{-1}$. This corresponds with the finding in a previous experimental study of PTFE amorphous polymer deposition on fluoropolymer thin films [75]. Similarly, as in the case of PDMS, the same peaks appear in the spectra of the PTFE bulk and surface models, with the main difference being in their intensity. Two dominant peaks of PTFE spectra are associated with symmetric and asymmetric stretching motions of the CF_2 bond, occurring at $1,155$ and $1,220\text{ cm}^{-1}$ according to experimental results [75–79]. The calculated values are $1,226$ and $1,248\text{ cm}^{-1}$ for surface and bulk PTFE models, and $1,135$ and $1,136\text{ cm}^{-1}$, which are in very good agreement with experimental results. The difference is slightly larger in the case of the surface PTFE model than for the bulk model. One more important peak appears in both spectra ($1,183\text{ cm}^{-1}$ for the surface model and $1,174\text{ cm}^{-1}$ for the bulk PTFE model), which is more significant in the case of the bulk model. It also belongs to C–F bond stretching. The opposite is true for peaks in the frequency interval from 750 cm^{-1} to $1,100\text{ cm}^{-1}$, for which the intensity of all peaks is much larger for the surface PTFE model compared with the bulk spectrum. All the bands in this range can be assigned to C–C bond changes. These absorptions ($730\text{--}750\text{ cm}^{-1}$) are known to arise in PTFE with amorphous character [78]. The computationally obtained peaks (bulk PTFE model, 650 cm^{-1} ; surface PTFE model, 639 cm^{-1}) agree well with the experimental spectrum of PTFE film [75] and can be assigned to C–F bond bending. In summary, the theoretical and experimental IR spectra for PTFE models was generally consistent. However, as in the case of PDMS, the spectrum obtained for the bulk PTFE model agrees slightly better with the experimental data than the results calculated for the surface PTFE model.

Surface energy

The goal of the present calculations on the studied polymers is to estimate the surface energies. These were calculated directly from the total energies of the surface and

bulk models after their ab initio periodic structure relaxation by applying the formula described in Eq. 1.

First, we would like to discuss changes in the total energies of both polymers over the time of calculation. Plots of time evolution of total energy trajectory versus time for the energy relaxation of the bulk models of PDMS and PTFE are shown in Figs. 6 and 7. They reveal a large decrease over the first 200 fs for both PTFE and PDMS, followed by a less drastic lowering of the energies of the studied bulk model over the next 300 fs for the PTFE trajectory. In the case of PDMS, total energy also does not change much over next 300 fs. At about 500 fs for PTFE and 200 fs for PDMS, polymers achieve first level or short-timescale equilibrium. We suggest that the initial conformation is energetically unstable and this instability is tied to break up of bonds in the simulation box.

After this point, the energy continues to decline gradually and, after 1,500 fs, it reaches more-or-less steady values equal to about -293.5857 eV for bulk PTFE and -527.0994 eV for bulk PDMS. This means that second level or long-timescale equilibrium was achieved. This shows that the conformational irregularities present in the initial structures have been removed (second level of optimized structure is obtained within 1,300 fs of the first level minima for PDMS, and within 1,000 fs for PTFE). As one can see, the changes in energy of the PTFE model are more drastic than those for the bulk of PDMS. This corresponds with changes in geometrical parameters and results obtained from DOS analysis. Thus, we conclude that a 2,000-fs time frame should be sufficient to obtain the energy minima for the studied polymers.

The calculated Ese values are 0.0222 J m^{-2} for PDMS and 0.0367 J m^{-2} for PTFE. The range of experimental and theoretical surface energy values for the most common polymers has already been summarized [38, 41]. According to this published data, Ese values range from 0.0157 to 0.024 J m^{-2} for PDMS and from 0.018 to 0.0265 J m^{-2} for PTFE. As one can see, our theoretical Ese value for PDMS is in good agreement with the experimentally reported

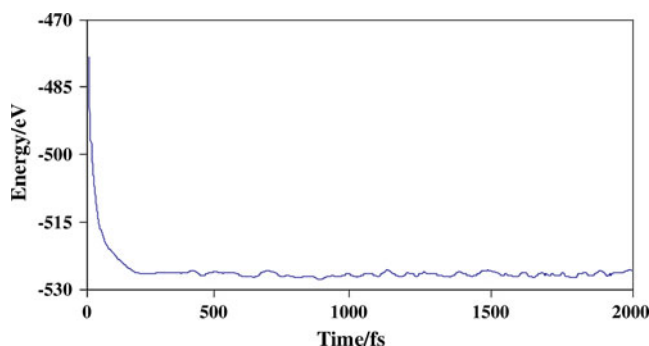


Fig. 6 Time evolution of the total energy trajectory versus time for the bulk model of PDMS during relaxation

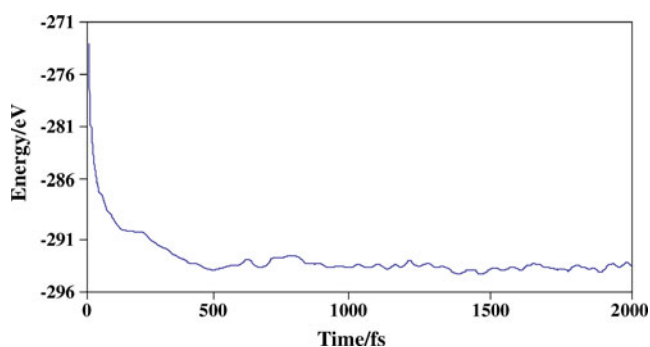


Fig. 7 Time evolution of the total energy trajectory versus time for the bulk model of PTFE during relaxation

results. However, in the case of PTFE, the calculated Ese value is almost two times larger.

One reason for this difference in the agreement of the obtained Ese values for PDMS and PTFE compared with the experimental data could be the different chemistry of these two coatings. The presence of alternating silicon and oxygen atoms in the backbone chain of polysiloxanes is fundamental to their properties. The Si–O bond length (1.64 Å, see Table 1 for more details) in PDMS is much longer than the typical C–C bond length (1.53 Å, see Table 2 for more details) in PTFE. Moreover, due to the lack of side groups on the oxygen, a relatively open intramolecular structure of PDMS is created compared to PTFE (carbon-based polymer). Such open structure strongly affects bending flexibility, conformational energetics (the conformational flexibility of PDMS is well-known) and backbone geometry of PDMS as well as the density and surface energy compared to carbon-based polymers. An important role in the unique characteristics of PDMS may be also played by the highly labile Si–O–Si angle. These large differences in their structures (two times higher density of PTFE than PDMS, higher molecular weight of PTFE than PDMS, which leads to different length of the chain per a unit cell) cause the geometrical features of the calculated models to vary significantly.

A similar structural dependency of surface and energetic properties has been reported in several publications. MD simulations of the water/trichloroethylene (TCE) interface show that the surface tension of long-tail systems is always lower than that of short-tail systems [80]. Application of this result to the PTFE models used in our study shows that, due to quite small models (short-tail), the obtained Ese should be higher than expected based on experimental values. The surface tension of polymers also exhibits a dependency on molecular weight [81]. This can be explained by the fact that, in the case of high molecular weight molecules (PTFE and PDMS belong to this category), the accessible surface area of such compounds is determined by the orientation of their segments, which is determined by the conformation of the molecules. A

quantum chemistry based force field study of PTFE reveals the molecular weight dependency of the relative energies only for certain studied conformers [22].

The surface energy reflects both molecular weight and density, as well as the character of the backbone. Thus, we conclude that bond flexibility and intramolecular interactions, including bond stretching, play a key role in the estimation of these values. Differences in the structures of the surface and bulk models after relaxation are much larger in the case of PTFE than PDMS (see above for more details on comparison of the geometrical and electronic parameters of PTFE and PDMS), which lead to much larger differences in total and Ese energies. Similar conclusions were drawn in a theoretical study on the amorphous PMMA–PTFE interface, where it was predicted that the PTFE layer undergoes greater deformation than the PMMA layer [39]. This behavior was accounted for by the energetics and conformational changes in the backbones, and entanglement in the bulk of the PTFE chain.

Another reason for the difference between calculated and experimental Ese values could be the limitations of the applied method. We suggest that this error is caused partially by using the LDA approach which has been found to overestimate the atomization energies for many systems (it fails in many cases to calculate correctly absolute values for binding energies, dissociation energies and weak intermolecular interactions because it does not include the dispersion forces). This suggestion is supported by the results of several other publications. It was shown that the absolute values of surface energies derived from DFT calculations depend on the applied approximations [82], and the Ese value should be corrected for the functional used [70, 83, 84]. The LDA and GGA-PBE applications were indicated to inaccurately describe electronic surfaces or edges [83] (differences in the calculated and experimental surface exchange energies are significant—almost two times higher using LDA and two times lower using PBE). This statement was confirmed in a theoretical study of atomic-scale properties of ZnO surfaces [85]. The authors of the latter study published a comparison of calculated cleavage energies of ZnO surfaces taken from the current literature [86–89], which revealed large differences in these energy values (in some case more than 100%). For example for (11–20) ZnO surface the energy value varies from 1.7 [89] to 4.1 J m⁻² [88] depending on the DFT functional used (B3LYP, LDA or GGA-PBE). A similar trend was found in our study but the difference is structure dependent.

Different substrate models

Due to the way in which unrelaxed models of the bulk and surface of PTFE and PDMS are prepared (algorithm in the

Amorphous builder [65] software and choice of cuts of the polymer structure to create the input simulation boxes, see [Substrate models](#) for more details), the initial structure is expected to play a significant role in the calculations.

Thus, after application of different cut offs of polymer structures to prepare the $[1 \times 1]$ unit cell of the surface and bulk PDMS and PTFE models, we analyzed how different initial structures and K-points influence E_{se} values. These new simulation boxes were prepared so that they contain different segments of the periodic structure of PTFE and PDMS. The unit cells possess the same dimensions and density but have a different percentage of atoms close to the borders of the cell or on the surface (mostly methyl group in PDMS and fluorine atoms in PTFE, see above for more details) than simulated in the original models. The new structures are less tangled together (more open), and have more broken bonds. The new calculations also included application of different K-points. The original calculations were performed using $1 \times 1 \times 1$ K-point mesh for both surface and bulk models. Then $2 \times 2 \times 2$ and $3 \times 3 \times 3$ K-point mesh were employed for the bulk models, and $2 \times 2 \times 1$ and $3 \times 3 \times 1$ for the surface models.

Relaxation of the new models resulted in far less compacted structures. The total energies of the bulk and surface models changed significantly upon application of different initial structures. These changes led to drastically higher E_{se} values for PDMS (more than 30 times higher). E_{se} value for PTFE was doubled. Changes in the applied grids resulted in different total energies of surface and bulk boxes, which caused an increase in E_{se} values (about seven times for PDMS and about three times for PTFE). This means that the equilibria and dynamics of the new fragments of PTFE and PDMS differ from those of the original cut offs. This also implies that PDMS model preparation is critical in these calculations, and that new PDMS models are not suitable to estimate E_{se} values correctly.

A similar dependency was found in several other publications. Substantially larger surface energy values were obtained in an MD study of PDMS under stress [90]. The surface energy of the monomer in the cavity is 230 mJ m^{-2} but experimental values range from 10 to 100 mJ m^{-2} [91]. The authors stated that the cause of such a large difference might be the relatively low density of the studied system. The surface tension of PDMS also increases as a function of chain length in the study of the dynamics of siloxane systems [29]. This finding is in agreement with the experimental results of Dee and Sauer [92, 93]. Other factors shown to influence the surface tension of PDMS are the end groups [29]—methyl-terminated PDMS have larger surface tension values than OH-terminated PDMS. Based on these results, and also on the fact that the surface tension of PDMS is very similar to that of alkane chains of

comparable length, the authors suggested that the methyl groups on the siloxane backbone play a greater role in determining the surface properties than does the backbone itself. Thus, different orientations and lengths of the S-CH₃ groups in our new models could lead to more significant differences in the total energies of the PDMS models than caused by various backbones (see the comparison of polysiloxane and carbon-based polymers in [Surface energy](#)).

The results obtained confirm the suggestion made above of a significant influence of the initial structure of the PTFE and PDMS surface and bulk models on the outcome of calculations. Thus, based on all previous comparisons, we conclude that the main cause of the large difference in E_{se} values seen using different fragments of the studied polymers is the bond cutting and their orientation in the surface region. The surface energy obtained from calculations of the surface enthalpy of solids from an ab initio electronegativity-based model also depends on the different directions applied to build the layers [94]. The number of bonds cut during the formation of a surface differs depending on the surface being formed. It is expected that the higher the number of bonds broken, the higher the surface energy per unit cell [95]. The calculated surface energy values follow this trend of increasing with the number of broken bonds, which is higher in the new models than in originally prepared simulation boxes with their less tangled structure.

Therefore, we conclude that there is no general rule for how to model the polymers to evaluate surface energy because differences in polymer structure critically influence various properties of these materials. Bond flexibility, angle and bond stretching, tail and backbone length (polymer shape) and functional groups near the surface of the model and their orientation affect the energetics of PDMS and PTFE greatly. We believe that, in calculations of surface energy performed using an MD approach with the plane-wave basis set, the most important factor is to pay attention to the configuration and conformity of the models used.

The calculations performed are limited to small models of PDMS and PTFE due to the obviously time-consuming calculations of the ab initio periodic structure relaxation. However, these first models appear sufficient due to the reasonably good agreement with experimental data at a level of accuracy that is unusual concerning the surface of solids [94].

Conclusions

PDMS and PTFE were investigated with a combined approach using classical MD and plane-wave pseudopotential methods. Calculations were performed using VASP.

Despite the existence of many publications on the surface energy of these two polymers, data obtained using various experimental or estimation techniques differ considerably. Thus, our main objective was to develop a modeling capability, using a combination of MD and molecular modeling methods, that would be able to replicate trends observed for substrates experimentally. It is of great interest to develop a fundamental understanding of the physics of the surface properties of selected coatings, and to develop the associated physically based models.

Full relaxation of models was allowed with periodic boundary conditions in all directions, and surface energies were calculated. The models and methods used were evaluated. The calculated surface energy values were validated through comparison of structural, electronic and energetic properties of PTFE and PDMS with experimental data [36–38, 41–44, 74–78] and other theoretical results [18–32, 72, 73, 96] to obtain the most realistic picture regarding the surface properties of the selected materials. The combination of MD calculations and use of specific bulk and surface models provides a new approach to obtaining features of PDMS and PTFE at the atomistic level and determining the surface energies.

Geometrical parameters as well as IR spectra for both polymers are in very good agreement with other theoretical studies. The calculated surface energy value for PDMS (0.022 J m^{-2}) is consistent with the experimental value. The surface energy of PTFE was larger (0.036 J m^{-2}) but still in reasonable agreement with the experimental results if one considers the errors that may occur due to the method used as discussed in the text. We concluded that this difference is caused mainly by the different structural nature of the studied polymers, including fundamentally different chain backbones, bond flexibility and lack of side groups on the oxygen. These differences result in quite an open intramolecular structure of PDMS compared to PTFE, which differentially affects calculation of the energetics of these polymers as well as surface energies. Moreover, the presence of different end groups near the borders of the simulation box also has a significant influence on these calculations. The method of bond cutting and different number of broken bonds in the models also plays a crucial role in determining the surface properties and energetics of these polymers.

Models of PDMS and PTFE were created to study the surface energy and surface properties of these two polymers, and were shown to provide a reasonable description. The method of calculations and modeling should be applicable to the calculation of such properties of other polymers.

Acknowledgment This work was funded by United Technologies Research Center.

References

- Burness JH, Dillard JG (1994) *Langmuir* 10:1894–1897
- Kim G, Ajersch F (1994) *J Mater Sci* 29:676–681
- Voronkov GM, Milileshevich VP, Yuzhelevich YA (1978) *The siloxane bond*. Consultants Bureau, New York
- Frischknecht AL, Curro JG (2003) *Macromolecules* 36:2122–2129
- Mark JE (1990) Silicon-containing polymers. In: Zeigler JM, Fearon FWG (eds) *Silicon-based polymer science; advances in chemistry series*. American Chemical Society, Washington, DC, p 47
- Zeigher JM, Fearon FWG (1990) *Silicon based polymer science: a comprehensive resource*, vol 224. ACS, Washington, DC
- Kraus G (1965) *Rubber Chem Technol* 38:1070–1114
- Vondracek AP (1990) *Rubber Chem Technol* 63:220–231
- Dee GT, Sauer BB (1994) *Macromolecules* 27:6106–6111
- Sauer BB, Dee GT (1994) *Macromolecules* 27:6112–6116
- Sakka T, Ogata YH (2005) *J Fluorine Chem* 126:371–375
- Hariharan A, Harris JG (1994) *J Chem Phys* 101:4156–4165
- Thomas R (1999) In: Hougham G, Johns K, Cassidy PE, Davidson T (eds) *Fluoropolymers 2: properties*. Plenum, New York
- Mansfield KF, Theodorou DN (1991) *Macromolecules* 24:6283–6294
- Chang J, Han J, Yang L, Jaffe RL, Yoon DY (2001) *J Chem Phys* 115:2831–2840
- Harris JG (1992) *J Phys Chem* 96:5077–5086
- Hapke T, Pätzold G, Heermann DW (1998) *J Chem Phys* 109:10075–10081
- Jang SS, Blanco M, Goddard WA III, Caldwell G, Ross RB (2003) *Macromolecules* 36:5331–5341
- Watkins EK, Jorgensen WL (2001) *J Phys Chem A* 105:4118–4125
- Collazo N, Shin S, Rice SA (1992) *J Chem Phys* 96:4735–4742
- Cui ST, Siepmann JI, Cochran HD, Cummings PT (1998) *Fluid Phase Equilib* 146:51–61
- Borodin O, Smith GD, Bedrov D (2002) *J Phys Chem B* 106:9912–9922
- Holt DB, Farmer BL, Macturk KS, Eby RK (1996) *Polymer* 37:1847–1855
- Lee S, Chang J, Jaffe RL, Yoon DY (2007) *Macromolecules* 40:7407–7412
- Sun H (1995) *Macromolecules* 28:701–712
- Smith JS, Borodin O, Smith GD (2004) *J Phys Chem B* 108:20340–20350
- Striolo A, McCabe C, Cummings PT (2005) *J Phys Chem B* 109:14300–14307
- Fritz L, Hofmann D (1997) *Polymer* 38:1035–1045
- Ismail AE, Grest GS, Heine DR, Stevens MJ (2009) *Macromolecules* 42:3186–3194
- Nath SK, Frischknecht AL, Curro JG, McCoy JD (2005) *Macromolecules* 38:8562–8573
- Bytner OG, Smith GD (2000) *Macromolecules* 33:4264–4270
- Macturk KS, Farmer BL, Eby RK (1995) *Polym Int* 37:157–164
- Bhattacharya S, Datta A, Berg JM, Gangopadhyay S (2005) *J Microelectromechanical Systems* 14:590–597
- Kroner E, Maboudian R, Arzt E (2010) *Adv Eng Mater* 12:398–404
- Hooper JB, Bedrov D, Smith GD, Hanson B, Borodin O, Dattelbaum DM, Kober EM (2009) *J Chem Phys* 130:144904–144911
- Choo BK, Song NY, Kim KH, Choi JS, Park KC, Jang J (2008) *J Non-Cryst Solids* 354:2879–2884
- Voue M, Semal S, De Coninck J (1999) *Langmuir* 15:7855–7862

38. Surface energy data for polydimethylsiloxane (PDMS) http://www.accudynetest.com/polymer_surface_data/polydimethylsiloxane.pdf
39. Okada O, Oka K, Kuwajima S, Toyoda S, Tanabe K (2000) *Comput Theor Polymer Sci* 10:371–381
40. Henry DJ, Yiapanis G, Evans E, Yarovsky I (2005) *J Phys Chem B* 109:17224–17231
41. Surface energy data for polytetrafluoroethylene (PTFE) http://www.accudynetest.com/polymer_surface_data/ptfe.pdf
42. Grundke K, Augsburg A (2000) *J Adhes Sci Technol* 14:765–775
43. Szymczyk K, Janczuk B (2007) *Langmuir* 23:8740–8746
44. Clint JH, Wicks AC (2001) *Int J Adhes Adhes* 21:267–273
45. Smith JS, Borodin O, Smith GD, Kober EM (2007) *J Polym Sci B Polym Phys* 45:1599–1615
46. Kresse G, Furthmüller J (1996) *Phys Rev B* 54:11169–11186
47. Kresse G, Hafner J (1993) *Phys Rev B* 48:13115–13118
48. Kresse G, Furthmüller J (1996) *J Comput Mater Sci* 6:15–50
49. Jones RO, Gunnarsson O (1989) *Rev Mod Phys* 61:689–746
50. Blöchl PE (1994) *Phys Rev B* 50:17953–17979
51. Kresse G, Joubert D (1999) *Phys Rev B* 59:1758–1775
52. Perdew JP, Zunger A (1981) *Phys Rev B* 23:548–552
53. Perdew JP, Wang Y (1992) *Phys Rev B* 45:13244–13249
54. Kroll P, Hoffmann R (1999) *J Am Chem Soc* 121:4696–4703
55. Ambrosch-Draxl C, Majewski JA, Vogl P, Leising G, Abt R, Aichholzer KD (1995) *Synth Metals* 69:411–414
56. Preat J, Rodríguez-Ropero F, Torras J, Bertran O, Zanuy D, Alemán C (2010) *J Comput Chem* 31:1741–1751
57. Brocks G, Kelly PJ, Car R (1993) *Synth Metals* 55–57:4243–4248
58. Ando S, Ueda M (2002) *Synth Metals* 129:207–213
59. Aouchiche HA, Djennane S, Boucekkine A (2004) *Synth Metals* 140:127–133
60. Jørgensen M, Sommer-Larsen P, Norrman K, Krebs FC (2004) *Synth Metals* 142:121–125
61. Izumi S, Hara S, Kumagai T, Sakai S (2004) *Thin Solid Films* 467:253–260
62. Nose S (1984) *J Chem Phys* 81:511–519
63. Ferrario M, Ryckaert JP (1985) *Mol Phys* 54:587–603
64. Monkhorst HJ, Pack JD (1976) *Phys Rev* 13:5188–5192
65. Scienomics Sarl (2004–2009) Amorphous builder software. http://www.scienomics.com/Products/classical_simulation/Amorphous-builder.php
66. Weijermars R (1986) *Tectonophysics* 124:325–258
67. Silicones and silicon-containing polymers, ABCR-Catalog 1994/1995. <http://www.abcr.de>
68. Merkel TC, Freeman BD, Spontak RJ, He Z, Pinnau I, Meakin P (2002) *Science* 296:519–522
69. Merkel TC, He Z, Pinnau I, Freeman BD, Meakin P, Hill (2003) *Macromolecules* 36:6844–6855
70. Swart JCW, van Helden P, van Steen E (2007) *J Phys Chem C* 111:4998–5005
71. Hara S, Kumagai T, Izumi S, Sakai S (2004) Structural and mechanical properties of amorphous silicon: *Ab initio* and classical molecular dynamics study. In: Proceedings of Second Multi-scale Materials Modeling (MMM-II) Conference, 11–15 October 2004, <http://www.icf11.com/proceeding/EXTENDED/4753.pdf>
72. Jothimuthu P, Carroll A, Asgar A, Bhagat S, Lin G, Mark JE, Papautsky I (2009) *J Micromech Microeng* 19:045024–045032
73. Xua J, Huang XH, Zhou NL, Zhanga JS, Baoa JCh, Lu TH, Li C (2004) *Mater Lett* 58:1938–1942
74. Shimomura M, Okumoto H, Kaito A, Ueno K (1998) *Macromolecules* 31(21):7483–7487
75. Nason TC, Moore JA, Lu TM (1992) *Appl Phys Lett* 60:1866–1868
76. Zamkov MA, Conner RW, Dlott DD (2007) *J Phys Chem C* 111:10278–10284
77. Schrader B (1989) *Raman/infrared atlas of organic compounds*, 2nd edn. VCH, Weinheim
78. de Wilde W, de Mey G (1973) *Vacuum* 24:307–311
79. FTIR spectra of polymers. <http://www.ftir-polymers.com/soon.htm>
80. Shi WX, Guo HX (2010) *J Phys Chem B* 114:6365–6376
81. Li Ch, Choi P (2006) *J Phys Chem B* 110:6864–6870
82. Mattsson AE, Jennison DR (2002) *Surf Sci* 520:L611–L618
83. Mattsson AE, Kohn W (2001) *J Chem Phys* 115:3441–3443
84. Mattsson TR, Mattsson AE (2002) *Phys Rev B* 66:214110–214117
85. Diebold U, Vogel Koplitz L, Dulub O (2004) *Appl Surf Sci* 237:336–342
86. Wander A, Schedin F, Steadman P, Norris A, McGrath R, Turner TS, Thornton G, Harrison NM (2001) *Phys Rev Lett* 86:3811–3814
87. Wander A, Harrison NM (2000) *Surf Sci* 457:L342–L346
88. Wander A, Harrison NM (2000) *Surf Sci* 468:L851–L855
89. Meyer B, Marx D (2003) *Phys Rev B* 67(035403):1–11
90. Lacevic NM, Maxwell RS, Saab A, Gee RH (2006) *J Phys Chem B* 110:3588–3594
91. Castellano M, Conzatti L, Costa G, Falqui L, Turturro A, Valenti B, Negroni F (2005) *Polymer* 46:695–703
92. Sauer BB, Dee GT (1991) *Macromolecules* 24:2124–2126
93. Dee GT, Sauer BB (1993) *Macromolecules* 26:2771–2778
94. Douillard JM, Henry M (2003) *J Colloid Interface Sci* 263:554–561
95. Spencer MJS, Hung A, Snook IK, Yarovsky I (2002) *Surf Sci* 513:389–398
96. Avramov PV, Fedorov DG, Irle S, Kuzubov AA, Morokuma K (2009) *J Phys Chem C* 113:15964–15968

Elucidation of specific aspects of dielectric constants of conjugated organic compounds: a QSPR approach

Areum Lee · Daejin Kim · Kyung-Hyun Kim ·
Seung-Hoon Choi · Kihang Choi · Dong Hyun Jung

Received: 18 January 2011 / Accepted: 23 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract The characteristic aspects of dielectric constants of π -conjugated compounds are elucidated by a quantitative structure property relationship (QSPR) study. To develop a QSPR model, among 141 collected π -conjugated compounds, a subset of 116 compounds was used as the training set for the model building and the rest was used as the test set for the model validation. Statistical regression models using 396 molecular descriptors were generated based on the genetic function approximation algorithm. The predicted dielectric constants obtained by the best model are highly correlated with the experimental values (squared correlation coefficient R^2 of 0.93 and 0.97 for the training and test sets, respectively), while a previous prediction model for general organic molecules (Sild S, Karelson M (2002) J Chem Inf Comput Sci 42:360–367) is not valid for our collected π -conjugated organic compounds. It has been known that the dielectric constants of organic materials are largely influenced by orientational correlations of the constituent molecules. In general, hydrogen bonding is one of the most important intermolecular interactions affecting orientational correlation. In the case of π -conjugated compounds, however, π - π interaction could be another comparable interaction with the hydrogen bonding.

Keywords Conjugated organic compound · Dielectric constant · Orientational correlation · Prediction model · QSPR

Introduction

The dielectric constant is the relative static permittivity (ϵ) measured as the ratio of the capacitance of a dielectric material relative to that of a vacuum. As one of the bulk properties of a material, the dielectric constant is largely influenced by orientational correlations. In organic liquids, molecular orientational correlations are significantly affected by various intermolecular interactions; the representative intermolecular interactions include dipole-dipole, higher-order multipole-multipole interactions and hydrogen bonding and recently σ -hole bonding has also been suggested to be an important interaction in determining orientational correlations [1, 2]. The importance of orientational correlations has been demonstrated in hydrogen-bonded organic materials. Goldman and Joslin showed that hydrogen-bonded liquids tend to have high static dielectric constants because the orientational correlation is enhanced by intermolecular hydrogen bonds [3].

Recently, π -conjugated organic compounds are widely used as electronic device materials. For optimal performance of organic-based devices, the organic compounds are often required to have dielectric constants in a specific range. For example, dielectric materials, such as organic thin films or crystals, in nonlinear optical devices must have a high dielectric constant for low optical loss and large electro-optic activity [4–6]. On the other hand, materials for interlayer dielectrics in integrated circuit devices must have a low dielectric constant to separate conducting parts efficiently [7, 8]. Understanding of intermolecular interactions and prediction of dielectric constants based on the

A. Lee · D. Kim · K.-H. Kim · S.-H. Choi · D. H. Jung (✉)
Insilicotech Co. Ltd., A-1101, Kolontripolis, 210,
Geumgok-Dong, Bundang-Gu,
Seongnam-Shi, Gyeonggi-Do 463–943, Republic of Korea
e-mail: dhjung@insilicotech.co.kr

D. Kim · K. Choi
Department of Chemistry, Korea University,
1, Anam-dong 5-Ga,
Seongbuk-Gu, Seoul 136–701, Republic of Korea

molecular structures would provide insights into developing π -conjugated organic compounds as new electronic materials.

Quantitative structure property relationship (QSPR) is a useful method to develop statistical regression models correlating physicochemical properties with molecular descriptors calculated from molecular structures. This QSPR approach has been successfully applied to predict chemical, physical, biological, and medicinal properties of various materials [9]. QSPR models have also been developed to predict dielectric constants of general organic compounds [10–13], organic solvents [14] or polymers [15], and especially the model by Sild and Karelson successfully reproduces the experimental dielectric constants of general organic compounds with various kinds of functional groups [12]. There are, however, no QSPR models reported to have good prediction ability for the dielectric constants of π -conjugated compounds. This is presumably due to the difficulty in the prediction of π - π interactions strongly affecting the orientational correlations in π -conjugated compounds. In this study, we propose a new QSPR model which can predict the dielectric constants of π -conjugated compounds successfully.

Data set

Collecting the dielectric constant data, we classified an organic compound as the π -conjugated if it has at least one double-single-double bond linkage in its structure. Since the dielectric constant is sensitive to temperature, collecting the experimental data measured at a specific temperature is crucial for building a reliable QSPR model. The experimental dielectric constants of π -conjugated compounds measured at 293.0–293.9 K were collected from the references [16, 17]. After a redundancy analysis, 141 compounds were selected for the data set with dielectric constants from 2.098 to 45.750 (Table 1). Then, the data set was divided randomly into two subsets; the training set with 116 compounds for model building and the test set with 25 compounds for model validation. The compounds in the training and test sets are listed in Table 2 and Table 3, respectively.

Table 1 Distribution of experimental dielectric constants for the training set and the test set

	Range of the dielectric constant					Total
	1~10	11~20	21~30	31~40	41~50	
Training set	101	12	2	1	0	116
Test set	18	4	2	0	1	25
Sum	120	16	4	1	2	141

Table 2 Experimental and calculated dielectric constants and their residual values for the training set

No.	Compound	Dielectric constant		
		exp.	GFA	Residual
1	allylbenzene	2.630	3.899	-1.269
2	4-allyl-2-methoxyphenol	9.550	6.929	2.621
3	aniline	7.060	7.888	-0.828
4	benzaldehyde	17.850	15.321	2.529
5	benzene	2.283	1.623	0.660
6	benzeneethanol	12.310	10.948	1.362
7	benzenepropanol	11.970	10.612	1.358
8	benzylamine	5.180	6.836	-1.656
9	benzylethylamine	4.300	2.999	1.301
10	bromobenzene	5.450	5.417	0.033
11	benzyl bromide	6.658	5.341	1.317
12	3-bromotoluene	5.566	6.728	-1.162
13	4-bromotoluene	5.503	7.233	-1.730
14	butylbenzene	2.359	2.396	-0.037
15	tert-butylbenzene	2.359	2.287	0.072
16	butyl phenyl ether	3.734	3.852	-0.118
17	butyl thiophene-2-carboxylate	6.400	6.798	-0.398
18	3-chloroaniline	13.300	11.273	2.027
19	chlorobenzene	5.690	6.102	-0.412
20	2-chloro-1,3-butadiene	4.914	7.955	-3.041
21	thiophene	2.739	3.313	-0.574
22	toluene-2,4-diisocyanate	8.433	9.192	-0.759
23	L-Nicotine	8.937	9.348	-0.411
24	2-methylanisole	3.502	3.274	0.228
25	3-methylanisole	3.967	3.843	0.124
26	4-methylanisole	3.914	3.136	0.778
27	α -methylbenzenemethanol	8.770	7.384	1.386
28	phenyl pentanoate	4.300	3.843	0.457
29	phenyl propanoate	4.770	4.821	-0.051
30	1-phenyl-2-propanol	9.350	5.266	4.084
31	phenylhydrazine	7.150	5.707	1.443
32	phenyl isocyanate	8.940	8.259	0.681
33	octylbenzene	2.260	3.022	-0.762
34	2,3,4,5,6-pentachlorotoluene	4.800	5.822	-1.022
35	pentyl benzoate	5.070	6.738	-1.668
36	isopropenylbenzene	2.280	4.197	-1.917
37	isobutylbenzene	2.318	2.092	0.226
38	isopropylbenzene	2.381	2.289	0.092
39	1-methoxynaphthalene	4.020	3.787	0.233
40	ethylbenzene	2.446	2.937	-0.491
41	α -ethylbenzenemethanol	6.680	4.515	2.165
42	ethyl benzoate	6.200	4.965	1.235
43	hexachloro-1,3-butadiene	2.550	1.901	0.649
44	hexylbenzene	2.300	2.987	-0.687
45	hexyl benzoate	4.800	6.691	-1.891
46	N-ethylaniline	5.870	7.062	-1.192

Table 2 (continued)

No.	Compound	Dielectric constant		
		exp.	GFA	Residual
47	2-methyl-N,N-dimethylaniline	3.400	3.882	-0.482
48	4-methyl-N,N-dimethylaniline	3.900	2.942	0.958
49	6-methylquinoline	8.480	10.064	-1.584
50	8-methylquinoline	6.580	7.244	-0.664
51	iodobenzene	4.590	5.274	-0.684
52	<i>o</i> -dibromobenzene	7.860	7.402	0.458
53	<i>m</i> -dibromobenzene	4.810	4.997	-0.187
54	dibenzylamine	3.446	4.025	-0.579
55	dibenzyl ether	3.821	2.911	0.910
56	benzyl chloride	6.854	5.343	1.511
57	3-chlorophenol	6.255	7.293	-1.038
58	3-chlorotoluene	5.763	6.587	-0.824
59	4-chlorotoluene	6.250	7.291	-1.041
60	1,6-dimethylnaphthalene	2.725	2.566	0.159
61	fluorobenzene	5.465	5.941	-0.476
62	ethyl thiophene-2-carboxylate	6.180	6.852	-0.672
63	2-ethyltoluene	2.595	3.711	-1.116
64	3-ethyltoluene	2.365	2.776	-0.411
65	4-ethyltoluene	2.265	2.095	0.170
66	2-ethylpyridine	8.330	8.000	0.330
67	4-ethylpyridine	10.980	12.430	-1.450
68	ethyl 4-pyridinecarboxylate	8.950	9.800	-0.850
69	ethyl phenylacetate	5.320	4.527	0.793
70	ethoxybenzene	4.216	3.634	0.582
71	dimethyl phthalate	8.660	8.003	0.657
72	2,5-dimethylpyrazine	2.436	3.497	-1.061
73	2,4-dimethylpyridine	9.600	9.207	0.393
74	2,6-dimethylpyridine	7.330	6.694	0.636
75	dioctyl phthalate	5.220	6.568	-1.348
76	dipentyl phthalate	6.000	7.097	-1.097
77	3,5-dimethylanisole	3.711	3.034	0.677
78	2,3-dimethyl-1,3-butadiene	2.102	1.584	0.518
79	dibutyl phthalate	6.580	6.917	-0.337
80	<i>o</i> -dichlorobenzene	10.120	7.532	2.588
81	<i>m</i> -dichlorobenzene	5.020	6.095	-1.075
82	1,3-diethyl-5-methylbenzene	2.264	1.682	0.582
83	diethyl phthalate	7.860	7.267	0.593
84	<i>m</i> -Diethylbenzene	2.369	2.415	-0.046
85	<i>p</i> -Diethylbenzene	2.259	1.657	0.602
86	styrene	2.474	3.486	-1.013
87	2-vinylpyridine	9.126	8.104	1.022
88	4-vinylpyridine	10.500	10.021	0.479
89	<i>o</i> -xylene	2.562	3.397	-0.835
90	<i>m</i> -xylene	2.359	2.671	-0.312
91	phenyl butanoate	4.480	3.606	0.874
92	1,3,5-triethylbenzene	2.256	1.512	0.744
93	1,2,3-trimethylbenzene	2.656	3.404	-0.748

Table 2 (continued)

No.	Compound	Dielectric constant		
		exp.	GFA	Residual
94	1,2,4-trimethylbenzene	2.377	2.577	-0.200
95	1,3,5-trimethylbenzene	2.279	1.585	0.694
96	4-methylpyridine	12.200	10.278	1.922
97	nitrobenzene	35.600	31.133	4.467
98	<i>o</i> -nitrotoluene	26.260	26.474	-0.214
99	propylbenzene	2.370	2.983	-0.613
100	pyridine	13.260	11.372	1.888
101	quinoline	9.160	9.811	-0.651
102	amyl benzoate	5.100	4.905	0.195
103	anisaldehyde	15.800	16.093	-0.293
104	anisole	4.300	3.747	0.553
105	benzotrchloride	7.400	6.497	0.903
106	benzyl benzoate	4.800	5.930	-1.130
107	benzyl salicylate	4.100	6.696	-2.596
108	ethyl benzoylacetate	12.800	10.470	2.330
109	isoamyl salicylate	5.400	7.353	-1.953
110	isobutyl benzoate	5.900	4.497	1.403
111	<i>m</i> -nitrotoluene	23.800	28.297	-4.497
112	methyl benzoate	6.600	6.873	-0.273
113	methyl salicylate	9.000	6.882	2.118
114	phenyl acetate	6.900	9.589	-2.689
115	salicylaldehyde	13.900	15.849	-1.949
116	toluene	2.400	3.112	-0.712

Calculation methodologies

All the molecular structures were optimized by using the DMol³ program [18], a quantum mechanical code based on the density functional theory (DFT), in the Materials Studio 4.4[®] package. The Perdew, Burke, and Ernzerhof (PBE) exchange-correlation functional [19] was used in the calculations with the double numerical polarization basis set [20]. The partial charges were derived by fitting the electrostatic potential to an atom-centered point charge model. The principal axes of inertia of each optimized structure were analyzed and the principal moments of inertia I_1 , I_2 , and I_3 were aligned along the x-, y-, and z-axes, respectively. The I_3 is the principal axis with the highest moment of inertia of each molecule. The QSAR program in the Materials Studio 4.4[®] package was used to calculate 396 descriptors for each compound. These descriptors can be divided as fast, Jurs, spatial, molecular mechanical, semi-empirical, and quantum mechanical descriptors according to the methodologies employed in the calculation. Fast descriptors include topological descriptors and electrotopological state (E-state) keys. E-state keys are numerical values, computed for each atom in

Table 3 Experimental and calculated dielectric constants and their residual values for the test set

No.	Compound	Dielectric constant		Residual
		Exp.	GFA	
1	benzonitrile	25.900	19.749	6.151
2	benzoyl fluoride	22.700	15.977	6.723
3	3-bromoaniline	13.000	10.969	2.031
4	2-bromotoluene	4.641	5.947	-1.306
5	4-chloroanisole	7.840	6.068	1.772
6	1,2,3,4-tetrahydro-2-naphthol	11.700	5.366	6.334
7	2-nitroanisole	45.750	33.647	12.103
8	1-methylnaphthalene	2.915	2.625	0.290
9	pentyl cinnamate	4.890	6.623	-1.733
10	phenyl laurate	3.280	3.452	-0.172
11	heptylbenzene	2.260	2.713	-0.453
12	2-methyl-1,3-butadiene	2.098	4.363	-2.265
13	2-methylquinoline	7.240	6.898	0.342
14	4-methylquinoline	9.310	10.313	-1.003
15	2-methylpyridine	10.180	7.443	2.737
16	4-methyl-1,3-pentadiene	2.599	4.590	-1.991
17	2-chlorotoluene	4.721	5.971	-1.250
18	1-ethyl-3,5-dimethylbenzene	2.275	1.826	0.449
19	benzyl dichloride	6.900	7.280	-0.380
20	2,6-dimethylanisole	3.780	3.304	0.476
21	dihexyl phthalate	5.620	6.913	-1.293
22	<i>o</i> -diethylbenzene	2.594	3.288	-0.694
23	<i>p</i> -xylene	2.274	1.732	0.542
24	N-methylaniline	5.960	6.986	-1.026
25	benzyl alcohol	13.000	8.629	4.371

a molecule, which encode information about both the topological environment of that atom and the electronic interactions due to all other atoms in the molecule [21]. The topological part of the relationship is based on the graph distances between each atom and all other atoms in the molecule. The electronic part is based on an intrinsic state value, plus perturbation due to intrinsic state differences between atoms in the molecule. Jurs descriptors are calculated by mapping atomic partial charges onto solvent-accessible surface areas of individual atoms. These descriptors, therefore, combine shape and electronic information to characterize molecules. Fast descriptors, Jurs descriptors, and molecular mechanical descriptors were calculated based on the electrostatic potential charges of the compounds.

Results and discussion

There have been only a few reports on the successful prediction of the dielectric constants of organic compounds

by using QSPR methods. Before developing our own model, we tested the model by Sild and Karelson to correlate the predicted dielectric constants with the experimental values of the π -conjugated molecules in our data set. This reference model using six descriptors was developed based on 155 organic compounds, among which 37 compounds are also included in our own data set. We first performed the geometry optimization and then calculated the six descriptors of the π -conjugated compounds in our data set using the reported method [12]. The dielectric constants predicted for the 37 overlapping compounds showed good correlation with the experimental values (squared correlation coefficient R^2 of 0.9243). However, when applied to all 141 π -conjugated compounds, the reference model did not show good prediction ability ($R^2=0.5673$, Fig. 1).

In order to predict the dielectric constant of π -conjugated compounds more accurately, we developed a new QSPR model using genetic function approximation (GFA) algorithm [22]. The GFA algorithm approach has some advantages over other techniques. It builds multiple models rather than a single model by means of automatic selection in which descriptors to be used in the models are selected. The models are generated by three iterated operations of selection, crossover, and mutation of variables and are compared with respect to the scoring function to evaluate the quality of the models until the algorithm converges. In our model building process, R^2 was used as the scoring function and the minimum and maximum number of variables in each model is specified as 4 and 7, respectively. Ten of the top equations were returned and the best model is summarized by Eq. 1 with seven descriptors (Table 4). The uncertainties in the descriptor coefficients in Eq. 1 are shown in parentheses.

$$\begin{aligned} \text{predicted dielectric constant} = & 6.2720(0.284) \times X1 + 5.4598(1.379) \\ & \times X2 - 1.6199(0.371) \times X3 \\ & - 36.1469(2.060) \times X4 \\ & + 1.6799(0.399) \times X5 \\ & - 1.6118(0.028) \times X6 \\ & + 0.3294(0.041) \times X7 + 0.9141 \end{aligned} \quad (1)$$

The dielectric constants predicted by the model are correlated well with the experimental values ($R^2=0.9275$). The adjusted R^2 and the leave-one-out cross-validation coefficient (R^2_{CV}) are 0.9209 and 0.8932, respectively, and the square of standard error of prediction (S^2) for the model is 1.94. The large F-statistic ($F=192.24$) also indicates the model fit the data very well. Unlike other statistical results, the Friedman lack-of-fit (LOF) has been known to resist over-fitting, which is a problem often encountered in constructing regression models. The Friedman LOF value

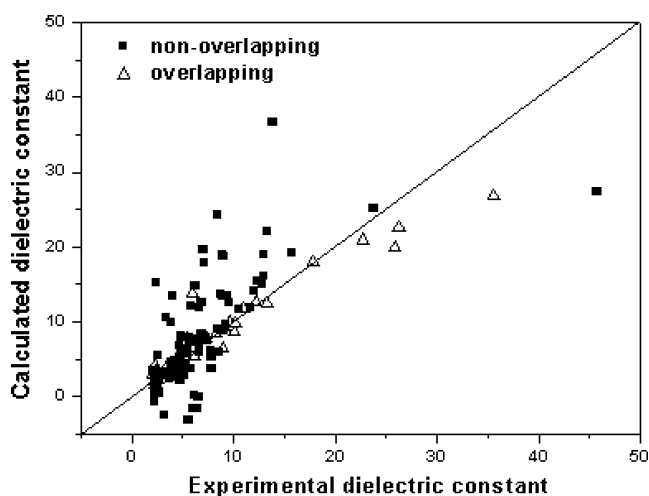


Fig. 1 Relationship between the experimental dielectric constants and the predicted results obtained by the reference QSPR model. Filled squares for the 141 non-overlapping compounds ($R^2=0.5673$) and empty triangles for the 37 overlapping compounds ($R^2=0.9243$)

for this model is 8.02. The average prediction error and the RMS error are 18.7% and 1.35 for the training set, and 23.7% and 3.51 for the test sets, respectively. The predicted and the experimental values of the test set also show significant correlation ($R^2=0.9696$, Fig. 2).

The descriptors in our best model consist of two quantum mechanical descriptors (total dipole, dipole z), four fast descriptors (S_{sssN} , S_{ddsN} , I_{aaN} , I_{ssO}), and one Jurs descriptor (RNCS). The dipole z is the z -vector component of the total dipole moment, where the z direction is coincident with the highest moment of inertia vector. The symbol S in descriptor name indicates the sum of the actual E-state keys of the atoms in specific type whereas the symbol I means the indicator for the presence (1) or the absence (0) of relevant atom. The symbol of sssN , ddsN , aaN , and ssO indicate the nitrogen atoms and the oxygen atom in conjunction with three single bonds, one single bond and two resonance single/double bonds as in nitro group, two aromatic bonds, and two single bonds, respectively. The plus and minus signs of t -test values in Table 4 show negative contributions of S_{sssN} , S_{ddsN} ,

Table 4 Descriptors used in Eq. 1

ID	t -test	Descriptor
X1	22.078	Total dipole
X2	3.959	Dipole z
X3	-4.369	E-state keys (sums): S_{sssN}
X4	-17.544	E-state keys (sums): S_{ddsN}
X5	4.208	E-state keys (indicators): I_{aaN}
X6	-5.689	E-state keys (indicators): I_{ssO}
X7	8.010	RNCS

and I_{ssO} to the dielectric constant and a positive contribution of I_{aaN} . The lone pair electrons in nitrogen or oxygen atoms can induce electronic repulsive intermolecular interactions which causes negative contributions to orientational correlations of constituent molecules. However, aromatic nitrogen atoms (aaN) can play a role in π - π attractive intermolecular interactions as well as in electronic repulsive interactions. Therefore, the positive t -test value of I_{aaN} can be explained as a result from dominant contributions of the aromatic nitrogens to π - π intermolecular interactions. The relative negative charged surface area (RNCS) is the solvent-accessible surface area of the most negative atom multiplied by the relative negative charge [23].

According to the t -test result (Table 4), the most important descriptor of our model is the total dipole moment, which is directly related to the dielectric constant as shown in the Debye's formula [24]. The t -test result also shows that the E-state keys are other important descriptors. Atoms with lone pair electrons tend to have large positive values of E-state keys and all four E-state keys of our model are related to nitrogen or oxygen atoms indicating the importance of the lone pair electrons of the constituent atoms.

Among the six descriptors of the reference model, the three most significant descriptors are the image of the Onsager-Kirkwood solvation energy, the number of hydrogen-acceptor sites and the fractional hydrogen bonding surface area; the image of the Onsager-Kirkwood solvation energy is the square of the total dipole moment of a molecule divided by the molecular weight and the other two descriptors quantify the strength of intermolecular hydrogen bonding [12]. Both in the reference and our models, the dipole moment is an important factor to predict the dielectric constant. However, whereas the descriptors

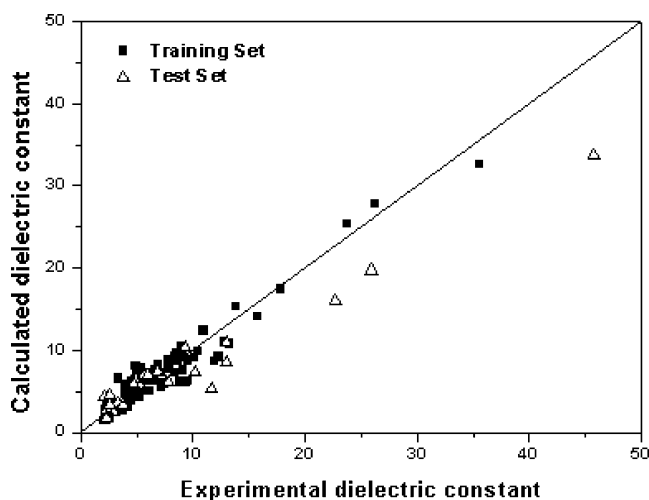


Fig. 2 Relationship between the experimental dielectric constants and the predicted results obtained by the newly developed QSPR model

related to hydrogen bonding are important in the reference model, they are not used as significant descriptors in our model. It is well known that hydrogen bonding can induce high orientational correlations [3]. Therefore the descriptors directly related to hydrogen bonding can have large positive contributions to the prediction of the dielectric constants of organic compounds in general. However, in the case of π -conjugated compounds, the π - π interaction can be the additional factor strongly affecting orientational correlation because the π - π interaction energy is comparable with the hydrogen bonding energy [25]. Because of more complicated intermolecular interactions between π -conjugated molecules, the contribution of hydrogen bonding should be treated in a different way for the better prediction of the dielectric constants of π -conjugated molecules.

Conclusions

The dielectric constants of 141 π -conjugated compounds have been modeled by a QSPR equation with seven descriptors. Compared to previous QSPR models, our model is more reliable in predicting the dielectric constants of π -conjugated molecules with RMS errors of 1.35 and 3.51 for the training and test sets, respectively. We expect that this new model could be applied to evaluate the dielectric constants of newly designed π -conjugated compounds and to prescreen a virtual library of electronic organic materials.

Acknowledgments This study was supported by grant No. 10031803 from the Industrial Source Technology Development Programs funded by the Ministry of Knowledge Economy, Republic of Korea. We thank Accelrys Korea for support with the modeling software.

References

1. Politzer P, Murray JS, Clark T (2010) Halogen bonding: an electrostatically-driven highly directional noncovalent interaction. *Phys Chem Chem Phys* 12:7748–7757
2. Murray JS, Riley KE, Politzer P, Clark T (2010) Directional weak intermolecular interaction: σ -Hole bonding. *Aust J Chem* 63:1598–1607
3. Goldman S, Joslin C (1993) Why hydrogen-bonded liquids tend to have high static dielectric constants. *J Phys Chem* 97:12349–12355
4. Raj CJ, Dinakaran S, Krishnan S, Milton B, Robert R, Das SJ (2008) Studies on optical, mechanical and transport properties of NLO active L-alanine formate single crystal grown by modified Sankaranarayanan–Ramasamy (SR) method. *Opt Commun* 281:2285–2290
5. Dhas SAMB, Natarajan S (2008) Growth and characterization of two new NLO materials from the amino acid family: L-Histidine nitrate and L-Cysteine tartrate monohydrate. *Opt Commun* 281:457–462
6. Lenin M, Ramasamy P (2008) Synthesis, growth and characterization of 3-nitroacetanilide—A new organic nonlinear optical crystal by Bridgman technique. *J Cryst Growth* 310:4451–4455
7. Zhao X, Wang M, Zhang B (2008) Structural and dielectric properties of conjugated polyacrylonitrile thin films deposited by plasma polymerization. *Thin Solid Films* 516:8272–8277
8. Dhara MG, Banerjee S (2010) Fluorinated high-performance polymers: Poly(arylene ether)s and aromatic polyimides containing trifluoromethyl groups. *Prog Polym Sci* 35:1022–1077
9. Katritzky AR, Maran U, Lobanov VS, Karelson M (2000) Structurally diverse quantitative structure-property relationship correlations of technologically relevant physical properties. *J Chem Inf Comput Sci* 40:1–18
10. Schweitzer RC, Morris JB (1999) The development of a quantitative structure property relationship (QSPR) for the prediction of dielectric constants using neural networks. *Anal Chim Acta* 384:285–303
11. Schweitzer RC, Morris JB (2000) Improved quantitative structure property relationships for the prediction of dielectric constants for a set of diverse compounds by Subsetting of the data set. *J Chem Inf Comput Sci* 40:1253–1261
12. Sild S, Karelson M (2002) A general QSPR treatment for dielectric constants of organic compounds. *J Chem Inf Comput Sci* 42:360–367
13. Liu J, Wilding WV, Giles NF, Rowley RL (2010) A quantitative structure property relation correlation of the dielectric constant for organic chemicals. *J Chem Eng Data* 55:41–45
14. Cocchi M, De Benedetti PG, Seeber R, Tassi L, Ulrici A (1999) Development of quantitative structure-property relationships using calculated descriptors for the prediction of the physicochemical properties (nD, ρ , bp, ϵ , η) of a series of organic solvents. *J Chem Inf Comput Sci* 39:1190–1203
15. Liu A, Wang X, Wang L, Wang H, Wang H (2007) Prediction of dielectric constants and glass transition temperatures of polymers by quantitative structure property relationships. *Euro P J* 43:989–995
16. Clipper Controls. Dielectric Constant Reference Guide. <http://clippercontrols.com/pages/dielectric-constant-values>. Accessed 18 Jan 2010
17. Lide DR (2000) CRC Handbook of Chemistry and Physics, 81st edn. CRC, Boca Raton
18. Delley B (2000) From molecules to solids with the DMol³ approach. *J Chem Phys* 113:7756–7764
19. Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. *Phys Rev Lett* 77:3865–3868
20. Delley B (1990) An all-electron numerical method for solving the local density functional for polyatomic molecules. *J Chem Phys* 92:508–517
21. Hall LH, Kier LB (1995) Electropotential State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *J Chem Inf Comput Sci* 35:1039–1045
22. Rogers D, Hopfinger AJ (1994) Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships. *J Chem Inf Comput Sci* 34:854–866
23. Stanton DT, Jurs PC (1990) Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies. *Anal Chem* 62:2323–2329
24. Onsager L (1936) Electric moments of molecules in liquids. *J Am Chem Soc* 58:1486–1493
25. Zhao Y, Truhlar DG (2008) Density functionals with broad applicability in chemistry. *Acc Chem Res* 41:157–167

Electron-topological, energetic and π -electron delocalization analysis of ketoenamine-enolimine tautomeric equilibrium

Agata Martyniak · Pawel Lipkowski · Noel Boens · Aleksander Filarowski

Received: 5 December 2010 / Accepted: 28 March 2011 / Published online: 27 April 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The ketoenamine-enolimine tautomeric equilibrium has been studied by the analysis of aromaticity and electron-topological parameters. The influence of substituents on the energy of the transition state and of the tautomeric forms has been investigated for different positions of chelate chain. The quantum theory of atoms in molecules method (QTAIM) has been applied to study changes in the electron-topological parameters of the molecule with respect to the tautomeric equilibrium in intramolecular hydrogen bond. Dependencies of the HOMA aromaticity index and electron density at the critical points defining aromaticity and electronic state of the chelate chain on the transition state (TS), OH and HN tautomeric forms have been obtained.

Keywords Aromaticity · Carbonylamine · Enolimine · Intramolecular hydrogen bond · QTAIM · Tautomeric equilibrium · HOMA

Introduction

The carbonylamines presented in the paper (Scheme 1) are prototypes of malonaldehydes investigated in a number of experimental and computational studies [1–3]. The given compounds can be referred to so-called resonance assisted hydrogen bond (RAHB) system being elaborated by Gilli et al. [4]. Recently the interest in RAHB has much grown due to the critics presented in refs [5, 6]. It is noteworthy that the aryl Schiff bases were properly studied [7–9]. However, theoretical studies of the carbonylamine (alkyl derivatives of Schiff bases) are not as numerous [10–14]. Both the alkyl Schiff bases and the aryl Schiff bases (Scheme 1) contain a quasi-aromatic formation as a common feature. But, the experimental data [15–23] expose the principle difference between these two types of compounds. What makes them different is that the HN (ketoenamine) tautomeric form prevails for the alkyl derivatives [10–26], meanwhile OH (enolimine) tautomeric form (Scheme 2) prevails for the aryl derivatives [7–9]. The transition from one form to another (the proton transfer) requires energetic expenditure (ΔE_{PT}). The change of tautomeric equilibrium is evoked by an acid-base balance and polarity of the environment. A series of papers [27–40] deals with the ways the tautomeric equilibrium affects the aromatic state of a molecule with intramolecular hydrogen bonding.

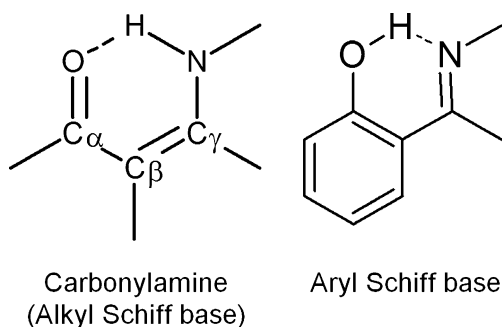
To trace a change of a tautomeric equilibrium in carbonylamines experimentally is really challenging. It is also difficult to estimate the substituent influence in the chelate chain ($O=C-C=C-N$) on the hydrogen bonding strength. This task is hampered by complementary phenomena (trans-cis isomerization [11] and $HO-CH=CH-C=NR \rightleftharpoons O=C-CH_2-CH=NR$ equilibrium [15–17]) observed experimentally for carbonylamines. However, quantum-

A. Martyniak · A. Filarowski (✉)
Faculty of Chemistry, University of Wrocław,
14 F. Joliot-Curie str.,
50-383 Wrocław, Poland
e-mail: afil@ruc.dk

A. Filarowski
e-mail: afil@wchuwr.chem.uni.wroc.pl

P. Lipkowski
Theoretical Chemistry Group, Institute of Physical and
Theoretical Chemistry, Wrocław University of Technology,
Wyb. Wyspiańskiego 27,
50-370 Wrocław, Poland

N. Boens
Department of Chemistry, Katholieke Universiteit Leuven,
Celestijnenlaan 200f – bus 02404,
3001 Leuven, Belgium

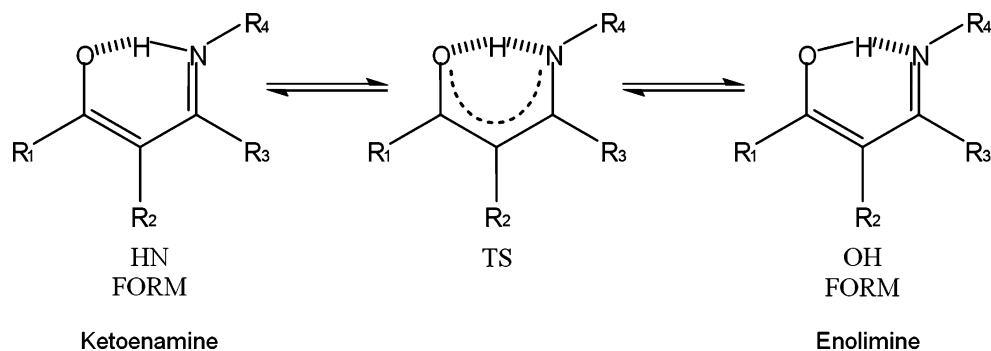


Scheme 1 Alkyl and aryl Schiff bases

mechanical calculations make it possible to describe a complicated nature of hydrogen bond. This paper is concerned with the ketoenamine-enolimine tautomeric equilibrium (Scheme 2) and the calculations of the carbonylamines with various substituents (CH₃, H, NH₂, Cl and F) in the chelate chain. The calculations were performed for the OH and HN tautomeric forms and transition state (Scheme 2) describing two main stages of the proton transfer process [41]. For description of these stages we calculated the difference of energies between them ($\Delta E_{PT}=E_{OH} - E_{HN}$, $\Delta E_{TS}=E_{TS} - E_{HN}$; the energy of HN tautomer is taken as the ground energetic level) and the HOMA aromaticity index [42] and electron-topological parameters [43, 44]. HOMA index is a widely accepted parameter applied in the description of the aromaticity [45, 46]. The merit of HOMA index is in its ability to be used for the experimental [47] and computational [48] results. The difference of energies is one of the most reliable parameters in the description of the hydrogen bonding [49].

As for the HOMA aromaticity index, it has been repeatedly verified by the researchers of intramolecular hydrogen bonding during last decade [26, 27, 30–40]. There are papers dwelling on the influence of tautomeric equilibrium on the aromaticity of phenol, aniline and naphthol complexes [50]. This paper develops these studies with aim to unify the HOMA aromaticity index in the area of hydrogen bonding and electron-topological analysis.

Scheme 2 Where R₁, R₂, R₃ and R₄ are CH₃, H, NH₂, Cl and F substituents



Computational details

The calculations were performed with Gaussian 03 [51] sets of code using the 6-311+G(d,p) basis set [52–55] at the Møller–Plesset second-order perturbation level (MP2) [56]. QTAIM analysis was performed using the AIM2000 program [57] with all the default options.

The HOMA aromaticity index was calculated by the following formula:

$$\text{HOMA} = 1 - \frac{1}{n} \sum_{i=1}^n \alpha_i (R_{opt} - R_i)^2 \quad (1)$$

where n is a number of bonds, α is an empirical constant, R_{opt} and R_i are the optimal and individual bond lengths taken from ref. 42. A higher HOMA aromaticity index corresponds to a more delocalized π -electronic system, hence a more aromatic formation [58].

Results and discussion

The influence of substituents on the intramolecular hydrogen bonding in the carbonylamines

The influence of substituents (R₁ – R₃) in the α , β and γ positions on the ketoenamine-enolimine equilibrium can be described by the inductive constants [59] ($\sigma_F=0$, ~ 0 , 0.14, 0.44 and 0.45 for H, CH₃, NH₂, Cl and F, respectively). It is noteworthy, that the π -electron donation of NH₂ group is small due to perpendicular orientation of its lone electron pair with respect to quasi-aromatic formation. This phenomenon was properly described by Sola et al. [60]. The increase of the electron acceptor ability of the R₁ and R₂ substituents (the increase of σ_F constant) in the α and β positions results in the growth of both the energetic barrier of transition state (ΔE_{TS}) and the OH form (ΔE_{PT}), (Fig. 1a and b). This trend is traced for the CH₃ (Fig. 1a) and H (Fig. 1b) substituents at the nitrogen atom, which serve as the basic ones for the predominant HN tautomeric form. An opposite picture is observed for the substituent (R₃) in the γ

Fig. 1 The energy levels of carbonylamine derivatives depending on substituents (black column respect to $R_1=H, CH_3, NH_2, Cl, F$; $R_2=R_3=H$; grey column respect to $R_2=H, CH_3, NH_2, Cl, F$; $R_1=R_3=H$; white column respect to $R_2=R_3=H$)

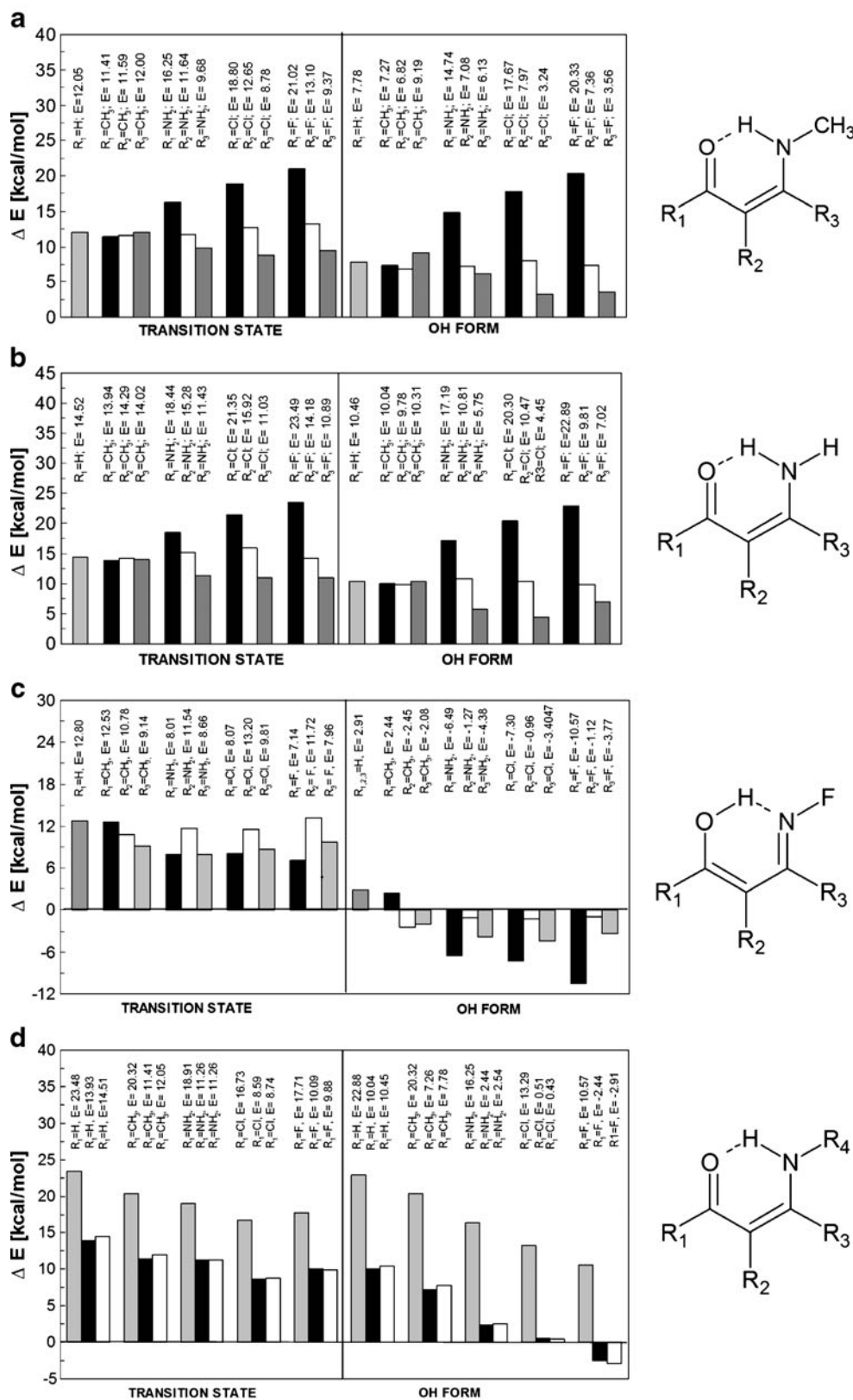
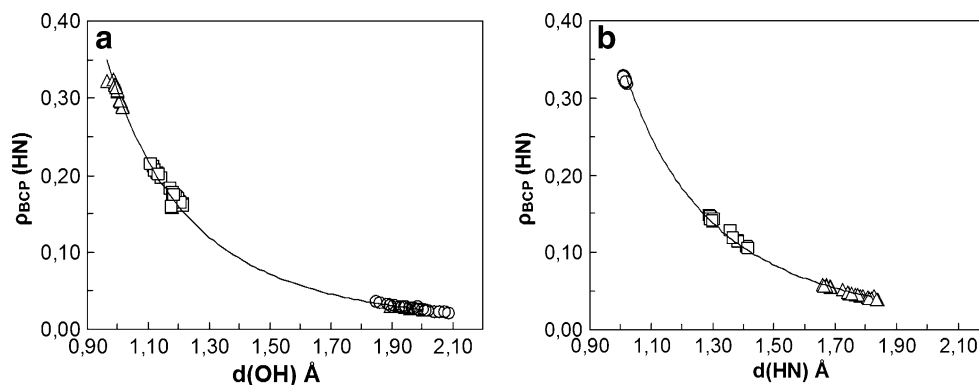


Fig. 2 Correlations between the electron density (**a**) at the OH bond critical point ($\rho_{\text{BCP}}(\text{OH})$, a.u.) and OH bond length ($d(\text{OH})$, Å), and (**b**) at the HN bond critical point ($\rho_{\text{BCP}}(\text{HN})$, a.u.) and HN bond length ($d(\text{HN})$, Å). Circles, squares and triangles correspond to the OH, transition state and HN forms, respectively



position (Fig. 1a and b) which reveals the decrease of the ΔE_{TS} and ΔE_{OH} values under the σ_{F} constants increase. These trends originate in the following phenomena: 1) the substitution (R_1) in the α position greatly affects the C=O group by weakening its basicity, consequently, it attenuates the hydrogen bond strength and enhances the ΔE_{TS} and $\Delta E_{\text{HN-OH}}$ barriers according to the CH_3 , H, NH_2 , Cl and F sequence; 2) the substitution (R_3) in the γ position mostly influences the amine group by increasing its acidity according to the Cl, F, NH_2 and $\text{H}\equiv\text{CH}_3$ sequence. Some exception from the expected CH_3 , H, NH_2 , Cl and F sequence is observed for the fluoro-substituent (R_3) in the γ position. The reason for this disagreement is a marked polarizability effect of the fluorine atom ($\sigma_{\alpha} = -0.25$ [59]) which causes some attenuation of the acidity of the HN group and the intramolecular hydrogen bonding. Remarkably, the substituent (R_2) in the β -position influences but to a lesser extent the ΔE_{TS} and ΔE_{OH} values due to its remote position from the acidic (NR_4) and basic ($\text{O}=\text{C}$) moieties.

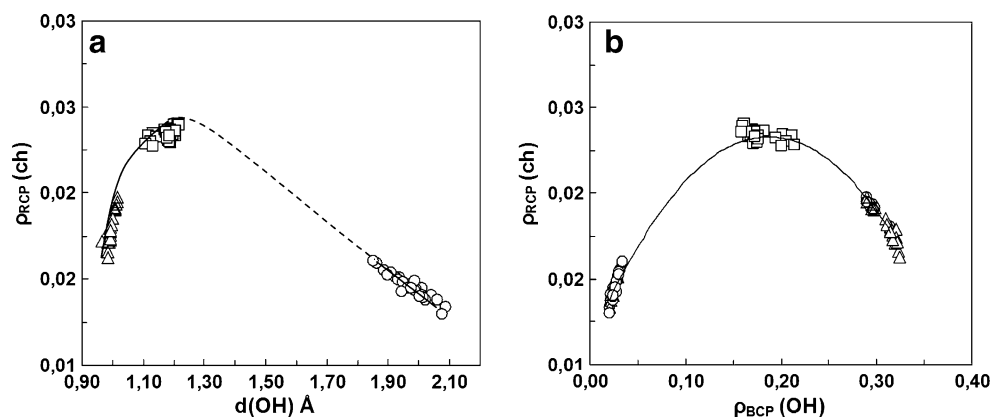
The picture changes for the N-F derivatives (R_4). The majority of these derivatives is characterized by the OH tautomeric form prevailing over the HN tautomeric form (Fig. 1c). The development of the electron acceptor ability of the R_1 substituent (under weak basicity of the nitrogen

atom, at N-F substituent) brings about both the decrease of the ΔE_{TS} values and the strengthening of the OH tautomeric form prevailing. However, the increase of the electron acceptor properties of the substituent (R_2) in the β -position is accompanied by the growth of the ΔE_{TS} values and the weakening of the OH tautomeric form prevailing.

With respect to the substituent impact on the nitrogen atom (R_4), the ΔE_{TS} and ΔE_{OH} values are getting smaller according to the H, CH_3 , NH_2 , Cl and F sequence (Fig. 1d). Some discrepancy as to the expected CH_3 , H, NH_2 , Cl and F sequence is observed for the H substituent which slightly influences the acidity of the amine group. A similar deviation was discovered for ortho-hydroxy aryl Schiff bases and explained by a significant polarization effect of the NH group [41, 60, 61].

In terms of the structural data of the hydrogen bridge ($d(\text{OH})$, $d(\text{HN})$ and $d(\text{OH})$), they are characterized by the following tendencies: 1) the elongation of the HN bond results in the reduction of the hydrogen bond and the OH bond lengths; 2) the elongation of the OH bond also triggers the reduction of the hydrogen bond and the HN bond lengths; 3) the shortest hydrogen bridge is found for the transition state; 4) the position of the TS is more shifted toward the reagents ($d(\text{O-H})_{\text{TS}} < d(\text{N-H})_{\text{TS}}$ for prevailing of the $\text{NH}\cdots\text{O}$ form) according to Leffler-Hammond rule [62, 63].

Fig. 3 Correlations between electron density at the chelate chain critical points ($\rho_{\text{RCP}}(\text{ch})$, a.u.) and (**a**) OH bond length ($d(\text{OH})$, Å), and (**b**) the electron density at the OH bond critical point ($\rho_{\text{BCP}}(\text{OH})$, a.u.)



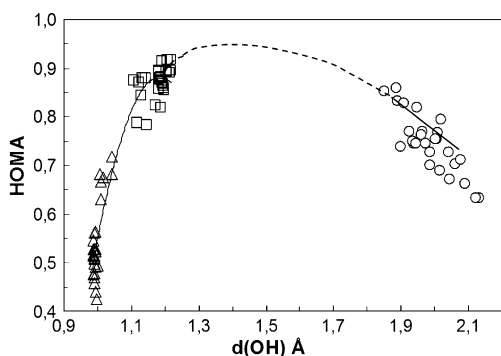


Fig. 4 Scatter plot of the chelate chain aromaticity index HOMA(ch) versus the OH bond length ($d(\text{OH})$, Å)

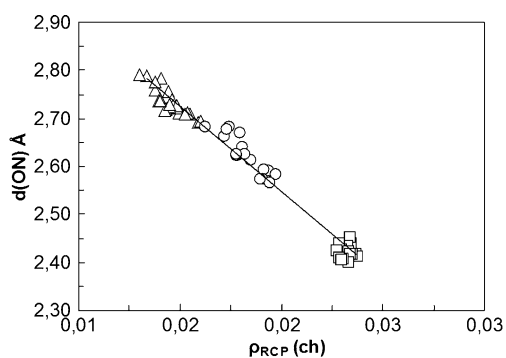
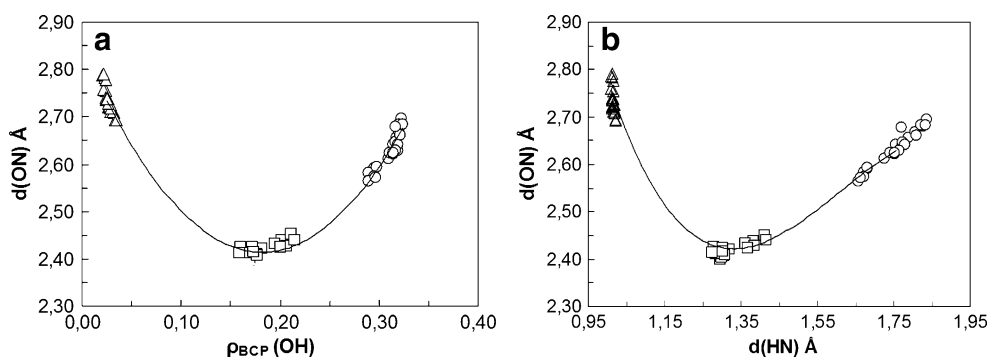
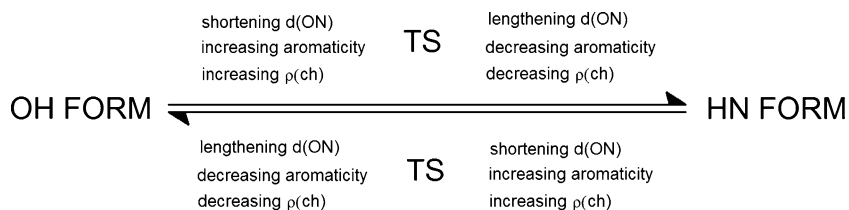


Fig. 5 Correlations between hydrogen bond length ($d(\text{ON})$, Å) and electron density at chelate chain critical points ($\rho_{\text{RCP}}(\text{ch})$, a.u.)

Fig. 6 Correlations between hydrogen bond length ($d(\text{ON})$, Å) and electron density at (a) chelate chain critical points ($\rho_{\text{RCP}}(\text{ch})$, a.u.), and (b) the HN bond length ($d(\text{HN})$, Å)



Scheme 3 Structural and electron-topological scheme of the tautomeric equilibrium



Analysis of aromatic and AIM parameters vs. tautomeric equilibrium

To verify the calculations obtained by the QTAIM method the $\rho_{\text{BCP}}(\text{XH})=f(d(\text{XH}))$ ($X=\text{N}$ or O atom) dependence was developed (Fig. 2; $\rho_{\text{BCP}}(\text{OH})=0.3489 \times (\text{OH})^{-3.522}$, $R^2=0.9991$; $\rho_{\text{BCP}}(\text{NH})=0.3071 \times (\text{NH})^{-3.5938}$, $R^2=0.9987$). The dependence taken by means of full optimization for different substituents in α - γ positions appears to be an exponential curve which is in accordance with the results obtained by non-adiabatic approach [41, 64]. Remarkably, the results for different substituents under full optimization are in agreement with those found under non-adiabatic approach for ortho-hydroxy aryl Schiff bases [65].

The next stage of the study was to verify the $\rho_{\text{RCP}}(\text{ch})=f(d(\text{OH}))$ and $\rho_{\text{RCP}}(\text{ch})=f(\rho_{\text{BCP}}(\text{OH}))$ dependencies (Fig. 3) calculated for the OH, TS and HN states of the compounds. These dependencies look like Morse or bell-shaped curves. A similar shape of the curve is traced for $\text{HOMA}(\text{ch})=f(d(\text{OH}))$ scatter plot (Fig. 4). The calculated scatter plot conditioned by some scattering of points supports the trends which are not observed under non-adiabatic approach [66]. The correlations reveal that the maximum of the electronic density at quasi-aromatic critical point ($\rho_{\text{RCP}}(\text{ch})$) and maximum aromaticity reaches its top in the transition state. This result comes in agreement with the results obtained in paper 66, where the comparison of the aromaticity of hydrogen-bonded and Li bonded aryl Schiff bases derivatives was carried out.

The $d(\text{ON})=f(\rho_{\text{RCP}}(\text{ch}))$ correlation (Fig. 5; $d(\text{ON}) = -35.482 \times (\rho_{\text{RCP}}(\text{ch})) + 3.2575$, $R^2=0.9849$) appears to be one of the most interesting dependencies which states that the shortening of the hydrogen bridge results from the growth of the electron density at the critical point of quasi-aromatic formation. But comparatively, the parabolic dependence $d(\text{ON})=f(\rho_{\text{BCP}}(\text{XH}))$ is more informative with respect to tautomeric equilibrium (Fig. 6) in terms of the basic parameter ($d(\text{ON})$) describing the hydrogen bond strength.

The $\rho_{\text{RCP}}(\text{ch})=f(d(\text{OH}))$, $\rho_{\text{RCP}}(\text{ch})=f(\rho_{\text{BCP}}(\text{OH}))$, $\text{HOMA}(\text{ch})=f(d(\text{XH}))$ and $d(\text{ON})=f(\rho_{\text{RCP}}(\text{OH}))$ dependencies (Figs. 3, 4 and 5) show that strengthening of the hydrogen bond brings about the increase of both electron density at the critical point of quasi-aromatic formation and the chelate chain aromaticity (Scheme 3).

Bearing in mind the fact of mutual increase of aromaticity and π -component in the aromatic formation one can confirm that the increase of $\rho_{\text{RCP}}(\text{ch})$ provokes the active participation of π -component.

Conclusions

It has been shown that for the HN tautomeric form the increase of the electron-acceptor ability of the substituents (R_1 and R_2) in α and β positions evokes a larger prevailing this form. In case of the OH tautomeric form the growth of the electron-acceptor ability of the substituents (R_1) in β position contributes into the OH tautomeric form prevailing, whereas for the substituent (R_2) in β position it reveals a reverse trend. With respect to the influence of substituent (R_3) in the γ -position for carbonylamines, the growth of the electron-acceptor ability hinders the prevailing of the HN tautomeric form. However, for the HN tautomeric form the impact of the substituent (R_3) in γ -position seems quite complicated due to mutual compensating action of the steric, inductive, resonance and polarizability effects, as well as local $\text{N-H}\cdots\text{F}$ hydrogen bond influence.

The $\text{HOMA}(\text{ch})=f(d(\text{OH}))$, $\rho_{\text{BCP}}(\text{ch})=f(d(\text{OH}))$, $d(\text{ON})=f(\rho_{\text{BCP}}(\text{ch}))$ dependencies have been obtained. The $\text{HOMA}(\text{ch})=f(d(\text{OH}))$ and $\rho_{\text{BCP}}(\text{ch})=f(d(\text{OH}))$ dependencies are bell-shaped and indicate that the transfer process from one tautomeric form to another goes via a transition state. According to these dependencies in the transition state one can observe the maximum delocalization of π -component of the chelate chain and the maximum electron density of quasi-aromatic formation. The $d(\text{ON})=f(\rho_{\text{BCP}}(\text{ch}))$ dependence states that the electron density at the critical points of quasi-aromatic formation is a measure of the intramolecular hydrogen bond strength. According to the presented dependencies the enhancement of the hydrogen bond strength leads to the growth of $\rho_{\text{BCP}}(\text{ch})$ and $\text{HOMA}(\text{ch})$ and reaches its maximum in the transition state. Krygowski et al. [58]

stated earlier that the increase of aromaticity strengthens π -component participation. Therefore, the strengthening of the hydrogen bond in the studied compounds is conditioned by the π -electron delocalisation in the chelate chain.

Acknowledgments The authors acknowledge the Wrocław Center for Networking and Supercomputing for generous computer time. P.L. acknowledges the Wrocław University of Technology for financial support.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Buemi G, Grabowski SJ (Eds) (2006) The hydrogen bonding—new insights. Springer, Berlin, pp 51–107
- Raczyńska ED, Kosińska W, Ośmiałowski B, Gawinecki (2005) Chem Rev 105:3561–3612
- Grabowski SJ, Leszczynski J (eds) (2006) The hydrogen bonding – new insights, Springer, Berlin, pp 487–512
- Gilli G, Gilli P (2009) The Nature of the Hydrogen Bond. Oxford University Press
- Sanz P, Mó O, Yanez M, Elguero J (2008) Chem Eur J 14:4225–4232
- Sanz P, Mó O, Yanez M, Elguero J (2009) Phys Chem Chem Phys 11:762–769
- Filarowski A, Koll A, Sobczyk L (2009) Curr Org Chem 13:172–193, and references cited herein
- Nedeltcheva D, Antonov L, Lycka A, Damyanova B, Popov S (2009) Curr Org Chem 13:217–239, and references cited herein
- Sheikhshoae I, Fabian WMF (2009) Curr Org Chem 13:149–171, and references cited herein
- Buemi G, Zuccarello F, Venuvanalingam P, Ramalingam M (2000) Theor Chem Acc 104:226–234
- Musin RN, Mariam YH (2006) J Phys Org Chem 19:425–444
- Bouchy A, Rinaldi D, Rivail J-L (2004) Int J Quantum Chem 96:237–281
- Lenain P, Mandado M, Mosquera RA, Bultinck P (2008) J Phys Chem A 112:10689–10696
- Raissi H, Moshfeghi E, Jalbout AF, Hosseini MS, Fazli M (2007) Int J Quantum Chem 107:1835–1845
- Kania L, Kamińska-Trela K, Witkowski M (1983) J Mol Struct 102:1–17
- Dąbrowski J, Kamińska-Trela K (1976) J Am Chem Soc 98:2826–2834
- Dąbrowski J, Kamińska-Trela K, Kania L (1976) Tetrahedron 32:1025–1029
- Dudek GO, Holm RH (1962) J Am Chem Soc 84:2691–2696
- Dudek GO, Dudek EP (1966) J Am Chem Soc 88:2407–2412
- Tayyari SF, Raissi H, Tayyari F (2002) Spectrochim Acta A 58:1681–1695
- Dudek GO, Holm RH (1961) J Am Chem Soc 83:2099–2104
- Brown NMD, Nouhebal DC (1968) Tetrahedron 24:5655–5664
- Martin R-F, Jamsosis GA, Martin BB (1961) J Am Chem Soc 83:73–75
- Rybarczyk-Pirek AJ, Grabowski SJ, Malecka M, Nawrot-Modranka J (2002) J Phys Chem A 106:11956–11962
- Nowroozi AR, Raissi H, Farzad F (2005) J Mol Struct THEOCHEM 730:161–169

26. Raczyńska ED (2005) *Pol J Chem* 79:1003–1006
27. Krygowski TM, Wozniak K, Anulewicz R, Pawlak D, Kolodziejski W, Grech E, Szady A (1997) *J Phys Chem* 101:9399–9404
28. Filarowski A, Koll A, Głowiak T, Majewski E, Dziembowska T (1998) *Ber Bunsenges Phys Chem* 102:393–402
29. Filarowski A, Koll A, Głowiak T (2002) *J Mol Struct* 615:97–108
30. Filarowski A, Kochel A, Cieślak K, Koll A (2005) *J Phys Org Chem* 18:986–993
31. Raczynska ED, Krygowski TM, Zachara JE, Osmialowski B, Gawinecki R (2005) *J Phys Org Chem* 18:892–897
32. Grabowski SJ, Sokalski WA, Leszczynski J (2006) *J Phys Chem A* 110:4772–4779
33. Grabowski SJ, Sokalski WA, Dyguta E, Leszczynski J (2006) *J Phys Chem B* 110:6444–6446
34. Palusiak M, Simon S, Sola M (2006) *J Org Chem* 110:5875–5822
35. Palusiak M, Krygowski TM (2007) *Chem Eur Chem* 13:7996–8006
36. Zubatyuk RI, Volovenko YM, Shishkin OV, Gorb L, Leszczynski J (2007) *J Org Chem* 72:725–735
37. Filarowski A, Kochel A, Kluba M, Kamounah F (2008) *J Phys Org Chem* 21:939–944
38. Kluba M, Lipkowski P, Filarowski A (2008) *Chem Phys Lett* 463:426–430
39. Karabiyik H, Petek H, Iskeleli NO, Albayrak C (2009) *Struct Chem* 20:903–910
40. Karabiyik H, Petek H, Iskeleli NO, Albayrak C (2009) *Struct Chem* 20:1055–1065
41. Filarowski A, Majerz I (2008) *J Phys Chem* 112:3119–3126
42. Krygowski TM (1993) *J Chem Inf Comput Sci* 33:70–78
43. Bader RF (1990) *Atoms in molecules: a quantum theory*. Oxford University Press, New York
44. Koch U, Popelier PL (1995) *J Phys Chem* 99:9747–9754
45. Krygowski TM, Cyrański MK (2001) *Chem Rev* 101:1385–1420
46. Sobczyk L, Grabowski SJ, Krygowski TM (2005) *Chem Rev* 105:3513–3560
47. Filarowski A, Kochel A, Cieslik K, Koll A (2005) *J Phys Org Chem* 18:986–993
48. Raczyńska ED, Krygowski TM, Zachara JE, Osmialowski B (2005) *J Phys Org Chem* 18:892–897
49. Korth H-G, de Heer MI, Mulder R (2002) *J Phys Chem A* 106:8776–8789
50. Krygowski TM, Zachara JE, Szatyłowicz H (2005) *J Phys Org Chem* 18:110–114, and references cited herein
51. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Lui G, Liashenko A, Piskosch P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) *Gaussian 03, Revision B.03*. Gaussian Inc, Wallingford, CT
52. McLean AD, Chandler GS (1980) *J Chem Phys* 72:5639–5648
53. Krishnan R, Binkley JS, Seeger R, Pople JA (1980) *J Chem Phys* 72:650–654
54. Clark T, Chandrasekhar J, Spitznagel GW, Schleyer PvR (1983) *J Comput Chem* 4:294–301
55. Frisch MJ, Pople JA, Binkley JS (1984) *J Chem Phys* 80:3265–3269
56. Møller C, Plesset MS (1934) *Phys Rev* 46:618–622
57. Biegler-König J, Schönbohm D, Bayles D (2001) *J Comput Chem* 22:545–559
58. Krygowski TM, Cyranski M, Czarnoski Z, Hafelinger G, Katritzky A (2000) *Tetrahedron* 56:1783–1796
59. Hansch C, Leo A, Taft RW (1991) *Chem Rev* 91:165–195
60. Fores M, Duran M, Sola M (2000) *Chem Phys* 260:53–64
61. Palomar J, De Paz JLG, Catalan J (2000) *J Phys Chem A* 104:6453–6463
62. Leffler JE (1953) *Science* 117:340–341
63. Hammond GS (1955) *J Am Chem Soc* 77:334–338
64. Espinosa E, Alkorta I, Elguero J, Molins E (2002) *J Chem Phys* 117:5529–5542
65. Filarowski A, Majerz I, Martyniak A (2010) *J Phys Chem A in preparation*
66. Krygowski TM, Zachara-Horeglad JE, Palusiak M (2010) *J Org Chem* 75:4944–4949

Microsolvation effect and hydrogen-bonding pattern of taurine-water TA-(H₂O)_n (*n*=1–3) complexes

Yumei Dai · Yuhua Wang · Zhengguo Huang ·
Hongke Wang · Lei Yu

Received: 31 December 2010 / Accepted: 24 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract The microsolvation of taurine (TA) with one, two or three water molecules was investigated by a density functional theory (DFT) approach. Quantum theory of atoms in molecules (QTAIM) analyses were employed to elucidate the hydrogen bond (H-bond) interaction characteristics in TA-(H₂O)_n (*n*=1–3) complexes. The results showed that the intramolecular H-bond formed between the hydroxyl and the N atom of TA are retained in most TA-(H₂O)_n (*n*=1–3) complexes, and are strengthened via cooperative effects among multiple H-bonds from *n*=1–3. A trend of proton transformation exists from the hydroxyl to the N atom, which finally results in the cleavage of the origin intramolecular H-bond and the formation of a new intramolecular H-bond between the amino and the O atom of TA. Therefore, the most stable TA-(H₂O)₃ complex becomes a zwitterionic complex rather than a neutral type. A many-body interaction analysis showed that the major contributors to the binding energies for complexes are the two-body energies, while three-body energies and relaxation energies make significant contributions to the binding energies for some complexes, whereas the four-body energies are too small to be significant.

Keywords Microsolvation · Hydrogen bond · Density functional theory · Quantum theory of atoms in molecules · Taurine · Many-body interaction analysis

Y. Dai · Z. Huang (✉) · H. Wang · L. Yu
Tianjin Key Laboratory of Structure and Performance for
Functional Molecules, College of Chemistry,
Tianjin Normal University,
Tianjin 300387, People's Republic of China
e-mail: hsxyhzg@126.com

Y. Wang
College of Materials Science and Engineering,
Shijiazhuang Tiedao University,
Shijiazhuang, Hebei 050043, People's Republic of China

Introduction

Taurine (TA), also known as 2-amino-ethanesulfonic acid (NH₂-CH₂-CH₂-SO₃H), is a relevant biomolecule in a variety of physiological processes such as the development of the brain and other tissues [1], calcium modulation, and osmoregulation [2]. Taurine also acts as a neurotransmitter in the central nervous system, activating some of the same receptors as γ -aminobutyric acid (GABA) [3, 4]. In addition, along with methionine, cystine and cysteine, TA is a sulfur amino acid and is considered a metabolic end product of methionine and cysteine [5]. The biological activity of TA depends on its molecular shape. A conformational search using AM1 and subsequently the B3LYP as well as MPn methods showed that three intramolecular hydrogen bonds (H-bonds) established between the amino and the sulfonic groups (N-H \cdots O=S, O-H \cdots N and N-H \cdots O-H) were found in the seven lowest-energy structures [6]. Laser ablation molecular beam Fourier transform microwave spectroscopy (LA-MB-FTMW) research showed that the conformational preference of TA is stabilized by an intramolecular H-bond O-H \cdots N between the hydroxyl group and the nitrogen atom of the amino group [7], while the structure of TA in crystal is stabilized as a zwitterion (NH₃⁺ - CH₂ - CH₂ - SO₃⁻) [8, 9]. Topological analysis of the experimental electron density and a comparison with high-level theoretical gas-phase calculations showed that the crystal environment has a significant influence on the electronic configuration of the sulfonate moiety in TA, which is more delocalized in the crystal than in the gas phase; this crystal effect is due mainly to hydrogen bonding interactions [8]. Similar to the crystal environment, we can infer that the aqueous environment also has a significant influence on the structure of TA since H-bonds are formed easily between

TA and water molecules. Moreover, TA in aqueous solution is closer to that in vivo. However, to the best of our knowledge, no theoretical studies on the structure of TA in the aqueous solution have been reported. Therefore, the aim of this paper was to study the microsolvation of TA, which builds a bridge between the gas phase and the solution phase. Theoretical and computational studies for the microsolvation effect on TA can provide a fundamental understanding of solvent–solute interactions at the molecular level, including the structural, spectral, energetic and thermodynamic properties, which can be important in attempting structure prediction and will aid future experimental and theoretical studies.

Hydrogen bonding interactions have a considerable effect on the microscopic as well as the macroscopic properties of biomolecules, and play a fundamental role in the understanding and designing of processes of biological and environmental importance [10, 11]. For example, large biomolecules have many sites that can participate in hydrogen-bonding interactions, and the lowest-energy conformations are often stabilized by intramolecular H-bonds between these sites. In some cases, however, the lowest-energy conformation of the isolated molecule is not suited to forming the strongest intermolecular H-bonds, i.e., the energy ordering of conformations can depend on the environment. Similarly, hydrogen bonding interactions are the major interaction in solutions of TA and play an important role in the microsolvation of TA. However, hydrogen bonding interactions are weaker than chemical bonding interactions, and not all theoretical methods reliably describe H-bonds. In general, the MP2 method is reliable to describe hydrogen bonding interactions. However, it is too time-consuming to apply to larger biomolecular systems even with a medium-size basis set. Compared with MP2, density functional theory (DFT) is a cost-effective approach; however, conventional DFT cannot describe hydrogen bonding interactions well ([12, 13] and references therein). Recently, some new DFT methods (such as B2PLYP [14], M06L [15, 16] and ω B97XD [17]) have been developed to deal with hydrogen bonding and van der Waals interactions. Comprehensive reviews evaluating such DFT methods were published by Johnson [18] and Hobza [19]. Many studies have shown that these new DFT methods can give reliable results for a wide variety of weakly bonded systems [20–26]. Recently, Dannenberg et al. [27] reported a comparison of the behavior of functional/basis set combinations for hydrogen-bonded systems, and they suggested that ω B97XD combined with 6-311++G(d,p) is a better choice.

The aim of this work was to study the microsolvation effect of TA by interacting with one, two, and three water molecules through hydrogen bonding interactions.

Quantum theory of atoms in molecules (QTAIM) analyses were employed to elucidate the hydrogen bonding interactions characteristic of TA-(H₂O)_n ($n=1-3$) complexes. The stable structures, binding energies and contributions of many-body energies to the binding energies of TA are discussed.

Computational details

DFT calculations

All DFT calculations were performed with Gaussian09 [28]. In this study, the ω B97XD functional [17] with the 6-311++G(d,p) basis set [29, 30] was used. The ω B97XD functional includes empirical dispersion and can treat hydrogen bonding and van der Waals interactions better than conventional DFT. First, the geometries of the isolated TA and water monomers were fully optimized. The TA-(H₂O)_n ($n=1-3$) complexes were constructed starting from the most stable TA and water monomers. All complexes were also fully optimized at the same level. The counterpoise (CP) correction [31] was implemented in order to ensure that complexes and monomers are being computed with a consistent basis set. The harmonic vibrational frequencies were calculated with analytic second derivatives at the same level, which confirm the structures as minima.

QTAIM analyses

QTAIM analyses were implemented using software AIM2000 [32] with ω B97XD wave functions employing the 6-311++G(d,p) basis set to provide complementary information on the H-bond. QTAIM has proven to be a very useful tool in describing electron densities in H-bond systems [33–35]. According to QTAIM, the topological properties of the bond critical point (BCP) between the X-H (donor of H-bond, H-donor) and the Y (acceptor of H-bond, H-acceptor) group can be used to study the nature of the H-bond. Both electron density (ρ_b) and the Laplacian of the electron density ($\nabla^2\rho_b$) at the BCP are a good measure of the strength of the H-bond. Two quantitative criteria proposed by Koch and Popelier are usually used to characterize the strength of the H-bond [36]: the electron density (ρ_b) and its Laplacian ($\nabla^2\rho_b$) should fall within the range of 0.002–0.04 and 0.02–0.15 a.u., respectively. The criteria provide a basis to distinguish these interactions from van der Waals interactions and have proved valid for standard and nonconventional H-bonds. In addition, the ring and cage structures formed consequently by many H-bonds are characterized by ring critical point (RCP) and cage critical point (CCP), respectively.

Many-body interaction analysis

The two-, three- and four-body contributions to total binding energy were calculated using many-body analysis [37–41]. The decomposition of the total energy of the complex can be written as

$$\begin{aligned} \Delta E &= E(1234) - \{E_{TA} + 3E_W\} \\ &= \sum_{i=1}^4 E(i) - \{E_{TA} + 3E_W\} && \text{relaxation energy} \\ &\quad + \sum_{i=1}^3 \sum_{j>i}^4 \Delta^2 E(ij) && \text{two - body energy} \\ &\quad + \sum_{i=1}^2 \sum_{j>i}^3 \sum_{k>j}^4 \Delta^3 E(ijk) && \text{three - body energy} \\ &\quad + \Delta^4 E(1234) && \text{four - body energy} \end{aligned} \quad (1)$$

$$\begin{aligned} \Delta^4 E(1234) &= E(1234) - \{E(1) + E(2) + E(3) + E(4)\} - \{\Delta^2 E(12) + \Delta^2 E(13) + \Delta^2 E(14) + \Delta^2 E(23) + \Delta^2 E(24) + \Delta^2 E(34)\} \\ &\quad - \{\Delta^3 E(123) + \Delta^3 E(124) + \Delta^3 E(134) + \Delta^3 E(234)\} \end{aligned} \quad (4)$$

The total binding energy of the complex is the sum of relaxation energy and the n -body interaction energies.

Results and discussion

In this work, the structure of isolated TA stabilized by an intramolecular O1H1^{TA}...N^{TA} H-bond in gas phase was well reproduced at the ω B97XD/6-311++G(d,p) level. The optimized conformer of TA was presented in Fig. 1. As shown in Fig. 1, the TA molecule can offer several possible donor and acceptor sites to form H-bonds. As the H-donor, the hydroxyl and amino groups of TA can donate protons to form H-bonds; moreover, they also can act as H-acceptors. In addition, the methylenes of TA as H-donors can also form H-bonds with water molecule in a few complexes. For water molecules, the hydroxyls are strong H-donors, while the oxygen atom can act as an H-acceptor to form an H-bond. From a structural viewpoint, an intramolecular O1H1^{TA}...N^{TA} H-bond can be formed between the hydroxyl group and the nitrogen atom in the most stable conformation of the free TA molecule. In addition, the NH3^{TA}...O1^{TA} intramolecular H-bond formed via a transfer of proton from the amino group to the oxygen atom is another possible intramolecular H-bond, which will lead also to instability of TA.

Molecular graphs of TA-(H₂O)_{*n*} ($n=1-3$) complexes formed by hydrogen bonding interactions are shown in Figs. 2, 3 and 4, while the structural parameters (R_{X-H} , ΔR_{X-H} , $R_{H...Y}$ and $\angle X-H...Y$) and electron topological properties (ρ_b and $\nabla^2 \rho_b$) of QTAIM analyses are listed in

where $E(i)$, $E(ij)$, $E(ijk)$, $E(1234)$ are the BSSE-corrected energies of the various monomers, dimers, trimers and tetramer in the complex, E_{TA} and E_W are the energies of isolated TA and water molecules, respectively. The pairwise two-body interaction energies and higher three-body and four-body interaction energies with BSSE correction are defined as the following equations:

$$\Delta^2 E(ij) = E(ij) - \{E(i) + E(j)\} \quad (2)$$

$$\begin{aligned} \Delta^3 E(ijk) &= E(ijk) - \{E(i) + E(j) + E(k)\} \\ &\quad - \{\Delta^2 E(ij) + \Delta^2 E(ik) + \Delta^2 E(jk)\} \end{aligned} \quad (3)$$

Tables 1, 2 and 3. Vibrational frequency calculations show that all complexes have no imaginary frequencies and are stable structures.

TA-(H₂O)

Four 1:1 complexes of TA with water formed by hydrogen bonding interactions were found at the ω B97XD/6-311++G(d,p) level. As shown in Fig. 2, all H-bonds in TA-(H₂O) complexes can be characterized by the existence of BCPs between XH...Y. Intramolecular H-bonds can be found in all TA-(H₂O) complexes. The O1H1^{TA}...N^{TA} H-bond are retained in TA-1W-1, TA-1W-2 and TA-1W-3, while the intramolecular H-bond in TA-1W-4 is the NH3^{TA}...O1^{TA} instead of the O1H1^{TA}...N^{TA} H-bond. As shown in Table 1, the intramolec-

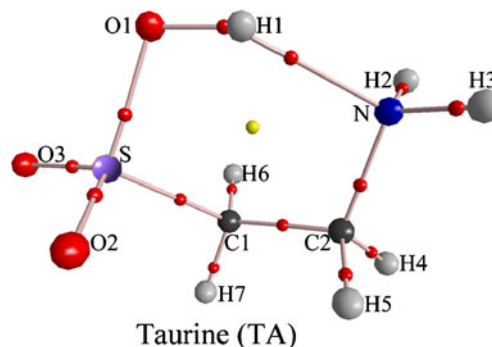
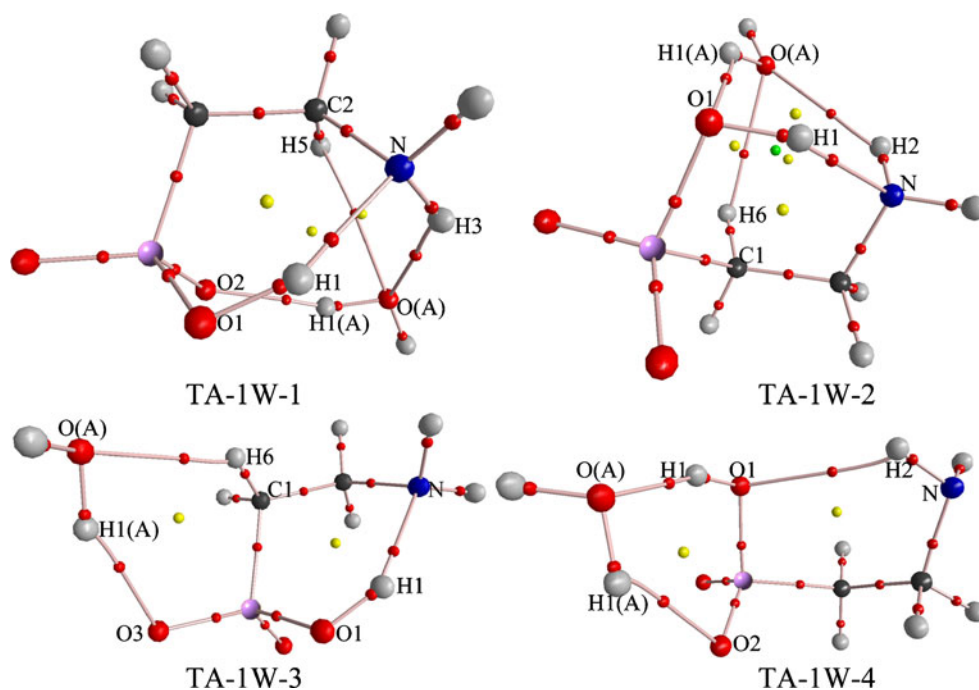


Fig. 1 Molecular graph of the monomer of free taurine (TA). *Large spheres* Attractors attributed to atomic positions: gray H, blue N, black C, red O, purple S. *Small spheres* Critical points: red bond critical point, yellow ring critical point

Fig. 2 Molecular graphs of TA-(H₂O) complexes. (A) Water molecules, *large spheres* attractors attributed to atomic positions: gray H, blue N, black C, red O, purple S. *Small spheres* Critical points: red bond critical point, yellow ring critical point, green cage critical point



ular $\text{NH}_3^{\text{TA}} \cdots \text{O}_1^{\text{TA}}$ H-bond in TA-1W-4 is a weak H-bond, being weaker than the $\text{O}_1\text{H}_1^{\text{TA}} \cdots \text{N}^{\text{TA}}$ H-bond in other TA-(H₂O) complexes, as can be seen from the unchanged $R_{\text{X-H}}$ and $R_{\text{H}\cdots\text{Y}}$ of 2.370 Å. As shown in Table 1, in three complexes except TA-1W-4, the intramolecular $\text{O}_1\text{H}_1^{\text{TA}} \cdots \text{N}^{\text{TA}}$ H-bonds involve positive $\Delta R_{\text{X-H}}$ values and the $R_{\text{H}\cdots\text{Y}}$ are shorter than that of the free TA molecule, which indicates that the intramolecular $\text{O}_1\text{H}_1^{\text{TA}} \cdots \text{N}^{\text{TA}}$ H-bond is strengthened when TA meets with one water molecule. Moreover, the order of intramolecular $\text{O}_1\text{H}_1^{\text{TA}} \cdots \text{N}^{\text{TA}}$ H-bond strength is:

TA-1W-1 > TA-1W-2 > TA-1W-3, since the $\Delta R_{\text{X-H}}$ value decreases and the H \cdots Y bond is elongated from TA-1W-1 to TA-1W-3. In addition, the results of QTAIM analyses show that the values of electron density (ρ_b) at the BCPs of H \cdots Y bonds of the intramolecular $\text{O}_1\text{H}_1^{\text{TA}} \cdots \text{N}^{\text{TA}}$ H-bonds are beyond the upper-limit values proposed by Koch and Popelier, which indicates that the intramolecular $\text{O}_1\text{H}_1^{\text{TA}} \cdots \text{N}^{\text{TA}}$ H-bonds in these complexes are strong H-bonds. Enhancement of the $\text{O}_1\text{H}_1^{\text{TA}} \cdots \text{N}^{\text{TA}}$ H-bonds is attributed to the cooperative effects that exist among multiple

Fig. 3 Molecular graphs of TA-(H₂O)₂ complexes. (A), (B) Water molecules; *large spheres* attractors attributed to atomic positions: gray H, blue N, black C, red O, purple S. *Small spheres* Critical points: red bond critical point, yellow ring critical point, green cage critical point

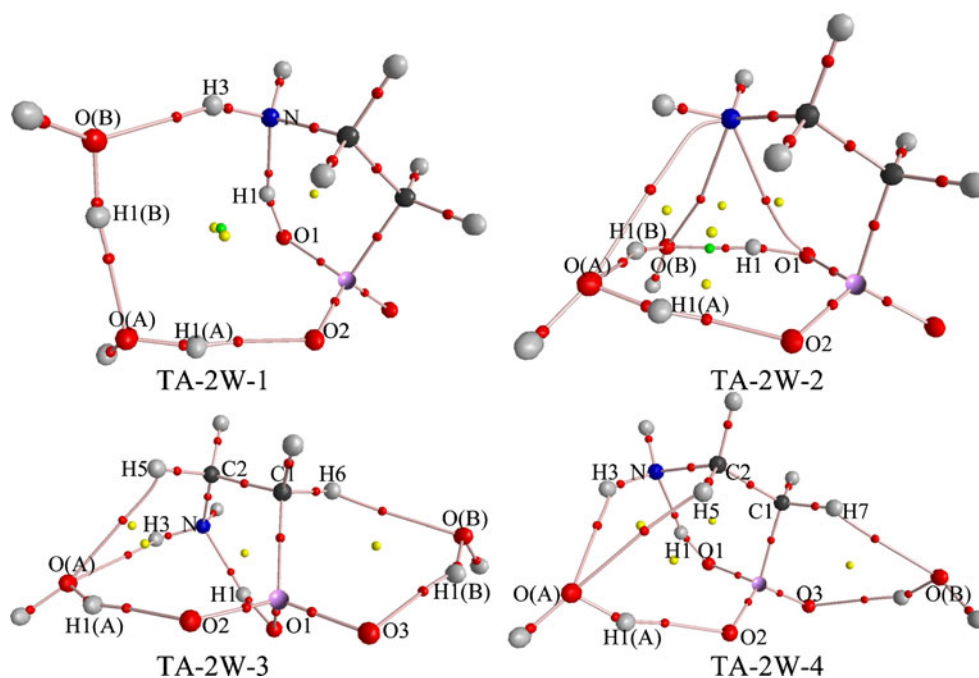
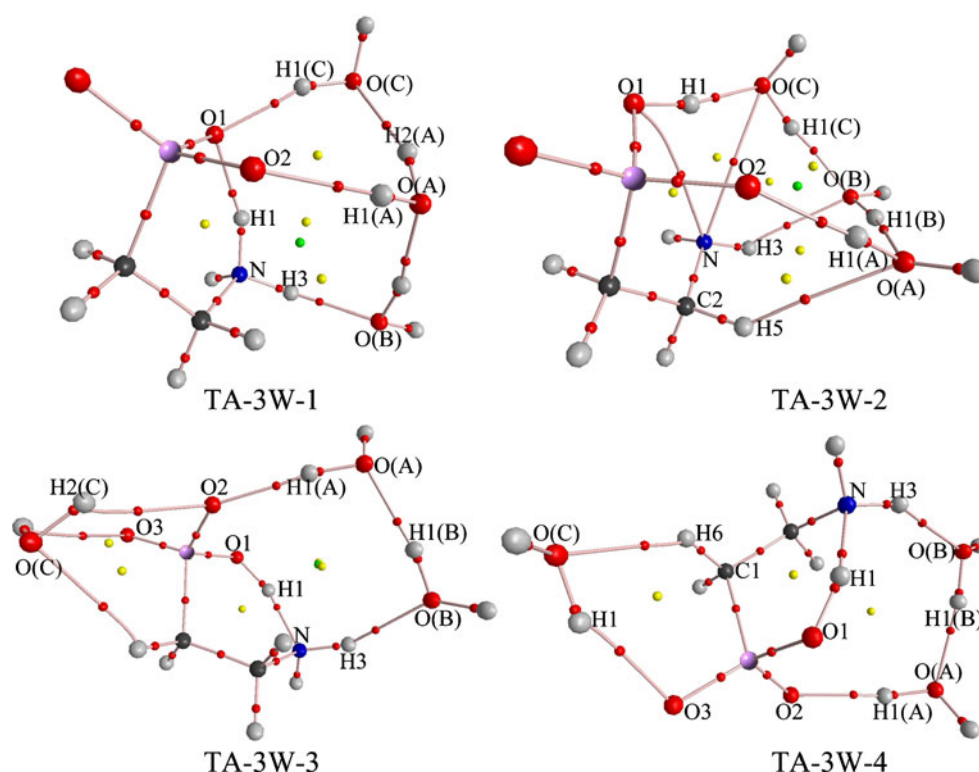


Fig. 4 Molecular graphs of TA-(H₂O)₃ complexes. (A), (B), (C) Water molecules; large spheres attractors attributed to atomic positions: gray H, blue N, black C, red O, purple S. Small spheres Critical points: red bond critical point, yellow ring critical point, green cage critical point



H-bonds in complexes. Cooperativity, i.e., the enhancement of the first H-bond between a H-donor and a H-acceptor when a second H-bond is formed between one of these two species and a third partner, is one of the hallmarks of hydrogen

bonding [42–44]. As shown in Fig. 2, for some complexes (TA-1W-1, TA-1W-2 and TA-1W-3), the N atom of TA becomes a stronger acceptor if the amino group donates an H-bond; moreover, the stronger donor leads to the stronger

Table 1 Structural parameters and electronic density topological analyses of hydrogen bonds (H-bonds) in taurine (TA)-(H₂O) complexes calculated at ω B97XD/6-311++G(d,p) level^a

Complex	H-bond	R_{X-H}	ΔR_{X-H}	$R_{H...Y}$	$\theta_{X-H...Y}$	ρ_b	$\nabla^2 \rho_b$
TA-1W-1	OH1 ^A ...O2 ^{TA}	0.970	0.012	1.883	159.0	0.0264	0.1050
	NH3 ^{TA} ...O ^A	1.018	0.006	2.077	134.3	0.0195	0.0751
	C2H5 ^{TA} ...O ^A	1.091	0.000	2.576	115.6	0.0078	0.0316
TA-1W-2	O1H1 ^{TA} ...N ^{TA}	1.012	0.019	1.710	153.5	0.0560	0.0994
	OH1 ^A ...O1 ^{TA}	0.965	0.007	2.085	136.8	0.0189	0.0704
	NH2 ^{TA} ...O ^A	1.018	0.004	2.280	128.2	0.0127	0.0473
	C1H6 ^{TA} ...O ^A	1.091	-0.001	2.469	128.2	0.0101	0.0327
TA-1W-3	O1H1 ^{TA} ...N ^{TA}	1.006	0.013	1.744	147.2	0.0513	0.1040
	OH1 ^A ...O3 ^{TA}	0.967	0.009	1.981	152.0	0.0224	0.0852
	C1H6 ^{TA} ...O ^A	1.093	0.001	2.332	137.0	0.0123	0.0413
TA-1W-4	O1H1 ^{TA} ...N ^{TA}	0.998	0.005	1.788	150.0	0.0462	0.1020
	O1H1 ^{TA} ...O ^A	0.992	-0.001	1.707	161.0	0.0436	0.1340
	OH1 ^A ...O2 ^{TA}	0.967	0.009	2.076	129.4	0.0199	0.0744
	NH3 ^{TA} ...O1 ^{TA}	1.012	0.000	2.370	127.5	0.0127	0.0418
TA	O1H1 ^{TA} ...N ^{TA}	0.993		1.828	148.2	0.0423	0.0996
	NH2	1.014					
	NH3	1.012					
	C1H6	1.092					
	C1H7	1.091					
	C2H5	1.091					
Water	OH	0.958					

^a Bond lengths in Ångstroms and bond angles in degrees, ρ_b and $\nabla^2 \rho_b$ are in a.u.

Table 2 Structural parameters and electronic density topological analyses of H-bonds in TA-(H₂O)₂ complexes calculated at ω B97XD/6-311++G(d,p) level^a

Complex	H-bond	R_{X-H}	ΔR_{X-H}	$R_{H...Y}$	$\theta_{X-H...Y}$	ρ_b	$\nabla^2\rho_b$
TA-2W-1	OH1 ^B ...O ^A	0.978	0.020	1.756	170.1	0.0377	0.1300
	OH1 ^A ...O2 ^{TA}	0.973	0.015	1.800	170.6	0.0317	0.1230
	NH3 ^{TA} ...O ^B	1.022	0.010	1.900	153.1	0.0274	0.1050
	O1H1 ^{TA} ...N ^{TA}	1.033	0.040	1.613	157.0	0.0704	0.0901
TA-2W-2	O1H1 ^{TA} ...O ^B	1.004	0.011	1.645	169.7	0.0505	0.1410
	OH1 ^B ...O ^A	0.975	0.017	1.837	161.0	0.0312	0.1110
	OH1 ^A ...O2 ^{TA}	0.975	0.017	1.793	166.9	0.0323	0.1250
TA-2W-3	OH1 ^A ...O2 ^{TA}	0.970	0.012	1.900	158.6	0.0255	0.1010
	OH1 ^B ...O3 ^{TA}	0.966	0.008	2.000	151.3	0.0216	0.0819
	NH3 ^{TA} ...O ^A	1.017	0.005	2.109	131.5	0.0183	0.0707
	C1H6 ^{TA} ...O ^B	1.093	0.001	2.365	135.7	0.0118	0.0386
	C2H5 ^{TA} ...O ^A	1.091	0.000	2.542	116.5	0.0083	0.0326
	O1H1 ^{TA} ...N ^{TA}	1.020	0.027	1.667	154.8	0.0618	0.0970
TA-2W-4	OH1 ^A ...O2 ^{TA}	0.970	0.012	1.897	158.5	0.0256	0.1020
	OH1 ^B ...O3 ^{TA}	0.966	0.008	2.011	148.4	0.0211	0.0804
	NH3 ^{TA} ...O ^A	1.018	0.006	2.084	133.7	0.0192	0.0742
	C1H7 ^{TA} ...O ^B	1.093	0.002	2.372	132.8	0.0117	0.0385
	C2H5 ^{TA} ...O ^A	1.090	-0.001	2.576	115.7	0.0078	0.0314
	O1H1 ^{TA} ...N ^{TA}	1.018	0.025	1.689	154.0	0.0589	0.0975

^a Bond lengths in Ångstroms and bond angles in degrees, ρ_b and $\nabla^2\rho_b$ are in a.u.

intramolecular O1H1^{TA}...N^{TA} H-bond in these complexes. Therefore, the order of cooperativity is TA-1W-1 > TA-1W-2 > TA-1W-3, which is consistent with the order of the strength of O1H1^{TA}...N^{TA} H-bonds in these complexes.

Except for the intramolecular H-bonds, multiple intermolecular H-bonds are formed in TA-(H₂O) complexes. As shown in Fig. 2, two H-bonds involving the amino and methylene of TA as the H-donors simultaneously can be found in TA-1W-1 and TA-1W-2, respectively. However, the C2H5^{TA}...O^A (TA-1W-1) and C1H6^{TA}...O^A (TA-1W-2) H-bonds are very weak, as can be seen from the almost unchanged ΔR_{X-H} and longer $R_{H...Y}$; moreover, the values of both ρ_b and $\nabla^2\rho_b$ are close to the lower-limit proposed by Koch and Popelier, which further confirm that these intermolecular H-bonds involving the methylene as H-donor are very weak, and partial van der Waals interaction characters can be attributed to them. In addition to two such H-bonds, one intermolecular H-bond is formed between the hydroxyl of water and the O of TA in TA-1W-1 and TA-1W-2, respectively. In either TA-1W-3 or TA-1W-4, two intermolecular H-bonds are formed, in which one H-bond involves the hydroxyl of water molecule as the H-donor while the other involves the oxygen of the water molecule as the H-acceptor. The intermolecular H-bond involving methylene as the H-donor in TA-1W-3 is slightly stronger than the corresponding bonds in either TA-1W-1 or TA-1W-2, while the intermolecular O1H1^{TA}...O^A H-bond in TA-1W-4 is the strongest H-bond among the four TA-(H₂O) complexes; its value of ρ_b (0.0436 a.u.) is beyond the upper-limit proposed by Koch and Popelier. In addition to

the intermolecular O1H1^{TA}...O^A H-bond in TA-1W-4, the other intermolecular H-bonds in TA-(H₂O) complexes are of intermediate strength; moreover, the H-bonds involving the hydroxyls of water as H-donors are stronger than those involving the amino of TA as H-donors, while the H-bonds involving the methylenes of TA as H-donors are the weakest.

TA-(H₂O)₂

Four 1:2 TA-(H₂O)₂ complexes formed by hydrogen bonding interactions were obtained at the ω B97XD/6-311++G(d,p) level. As shown in Fig. 3, the intramolecular O1H1^{TA}...N^{TA} H-bond is retained in all TA-(H₂O)₂ complexes except TA-2W-2, and the decrease in ΔR_{X-H} from TA-2W-1 to TA-2W-4 confirms the order of strength as TA-2W-1 > TA-2W-3 > TA-2W-4. By comparison with Tables 1 and 2, the intramolecular O1H1^{TA}...N^{TA} H-bonds in these TA-(H₂O)₂ complexes are stronger than those of TA-(H₂O) complexes since they involve larger ΔR_{X-H} and larger ρ_b as well as $\nabla^2\rho_b$. Moreover, the values of ρ_b at the BCPs of H1^{TA}...N^{TA} bonds are far beyond the upper-limit value proposed by Koch and Popelier. Similar to TA-(H₂O) complexes, the cooperative effect can be found in TA-(H₂O)₂ complexes as well. As shown in Fig. 3, for some complexes (TA-2W-1, TA-2W-3 and TA-2W-4), the formation of intermolecular H-bonds between the amino of TA and water molecules results in the N atom of TA becoming a stronger acceptor, which further strengthens the intramolecular O1H1^{TA}...N^{TA} H-bonds.

Table 3 Structural parameters and electronic density topological analyses of H-bonds in TA-(H₂O)₃ complexes calculated at ω B97XD/6-311++G(d,p) level^a

Complex	H-bond	R_{X-H}	ΔR_{X-H}	$R_{H...Y}$	$\theta_{X-H...Y}$	ρ_b	$\nabla^2\rho_b$
TA-3W-1	OH1 ^B ...O ^A	0.995	0.037	1.643	166.3	0.0514	0.1490
	NH3 ^{TA} ...O ^B	1.047	0.035	1.681	155.2	0.0472	0.1460
	OH1 ^C ...O1 ^{TA}	0.976	0.018	1.783	159.5	0.0344	0.1290
	OH2 ^A ...O ^C	0.974	0.016	1.834	160.0	0.0321	0.1130
	OH1 ^A ...O2 ^{TA}	0.974	0.016	1.860	161.4	0.0292	0.1080
	NH1 ^{TA} ...O1 ^{TA}	1.055		1.675	151.2	0.0511	0.1450
TA-3W-2	O1H1 ^{TA} ...O ^C	1.007	0.014	1.614	167.8	0.0539	0.1480
	OH1 ^B ...O ^A	0.979	0.021	1.744	163.4	0.0382	0.1340
	OH1 ^A ...O2 ^{TA}	0.976	0.018	1.738	175.6	0.0357	0.1390
	OH1 ^C ...O ^B	0.975	0.017	1.824	171.1	0.0316	0.1140
	NH3 ^{TA} ...O ^B	1.012	0.000	2.260	130.7	0.0136	0.0500
	C2H5 ^{TA} ...O ^A	1.092	0.001	2.607	142.3	0.0069	0.0245
TA-3W-3	OH1 ^B ...O ^A	0.978	0.020	1.759	169.6	0.0375	0.1300
	OH1 ^A ...O2 ^{TA}	0.973	0.015	1.803	171.8	0.0316	0.1220
	NH3 ^{TA} ...O ^B	1.023	0.011	1.894	152.5	0.0278	0.1070
	C1H7 ^{TA} ...O ^C	1.092	0.001	2.468	133.3	0.0107	0.0313
	OH2 ^C ...O2 ^{TA}	0.962	0.004	2.342	129.8	0.0126	0.0416
	OH1 ^C ...O3 ^{TA}	0.961	0.003	2.545	121.7	0.0090	0.0298
	O1H1 ^{TA} ...N ^{TA}	1.052	0.059	1.554	158.5	0.0814	0.0751
TA-3W-4	OH1 ^B ...O ^A	0.978	0.020	1.757	170.4	0.0376	0.1300
	OH1 ^A ...O2 ^{TA}	0.958	0.000	3.184	56.5	0.0313	0.1210
	NH3 ^{TA} ...O ^B	1.022	0.010	1.905	149.9	0.0272	0.1050
	OH1 ^C ...O3 ^{TA}	0.967	0.009	1.984	151.4	0.0223	0.0849
	C1H6 ^{TA} ...O ^C	1.093	0.001	2.325	140.3	0.0124	0.0415
	O1H1 ^{TA} ...N ^{TA}	1.061	0.068	1.523	159.4	0.0878	0.0653

^a Bond lengths in Ångstroms and bond angles in degrees, ρ_b and $\nabla^2\rho_b$ are in a.u.

Multiple intermolecular H-bonds can be found in TA-(H₂O)₂ complexes. For either TA-2W-1 or TA-2W-2, only one intermolecular H-bond is formed between one water molecule and TA. For TA-2W-3 or TA-2W-4, one water molecule forms two H-bonds with TA by involving both the amino and methylene of TA as the H-donors simultaneously, while another water molecule forms two intermolecular H-bonds with TA, in which the water molecules acts as the H-donor and H-acceptor, respectively. Because no intramolecular H-bond is formed in TA-1W-2, the cooperative effect occurring in TA-1W-2 enhances the strength of the intermolecular O1H1^{TA}...O^B H-bond, which is differ from what occurs in the other complexes, in which the cooperative effect strengthens the intramolecular O1H1^{TA}...N^{TA} H-bonds. Moreover, the intermolecular O1H1^{TA}...O^B H-bond in TA-1W-2 is the strongest intermolecular H-bond, involving the shortest $R_{H...Y}$ (1.645 Å) and the largest values of ρ_b (0.0505 a.u.) as well as $\nabla^2\rho_b$ (0.1410 a.u.). In addition, more cooperativity deriving from the intermolecular H-bonds formed between two water molecules in TA-2W-1 and TA-2W-2 are helpful in strengthening the intermolecular H-bonds, which are stronger than those in TA-2W-3 and TA-2W-4 since no H-

bonds are formed between the two water molecules in TA-2W-3 and TA-2W-4.

TA-(H₂O)₃

Four 1:3 complexes of TA with water molecules formed by hydrogen bonding interactions were found at the ω B97XD/6-311++G(d,p) level. As shown in Fig. 4, multiple H-bonds are formed between TA and water molecules in each complex. The intramolecular O1H1^{TA}...N^{TA} H-bond remained intact in both TA-3W-3 and TA-3W-4, while no intramolecular H-bond can be found in TA-3W-2. It is worth noting that TA-3W-1 becomes a zwitterionic complex because of the formation of the new intramolecular NH1^{TA}...O1^{TA} H-bond, in which the amino group is the H-donor rather than the H-acceptor. Even for TA-3W-3 and TA-3W-4, the trend of transformation from the intramolecular O1H1^{TA}...N^{TA} H-bond to the NH1^{TA}...O1^{TA} bond is very clear since the proton shifts to the middle site of the distance between the oxygen and nitrogen atom, as can be seen in Table 3. Moreover, the intramolecular O1H1^{TA}...N^{TA} H-bonds in both TA-3W-3 and TA-3W-4 are very strong since the values of ρ_b (0.0814 and 0.0878 a.

u.) at the BCPs of $\text{H1}^{\text{TA}}\cdots\text{N}^{\text{TA}}$ bonds are far beyond the upper-limit value proposed by Koch and Popelier, which also indicates the trend of transformation of the intramolecular $\text{O1H1}^{\text{TA}}\cdots\text{N}^{\text{TA}}$ H-bonds. The enhancement of intramolecular $\text{O1H1}^{\text{TA}}\cdots\text{N}^{\text{TA}}$ H-bonds is attributed to the cooperative effect due to the formation of the intermolecular H-bond between the amino of TA and a water molecule.

More intermolecular H-bonds are formed in $\text{TA}-(\text{H}_2\text{O})_3$ complexes and result in the formation of cage structures characterized by the CCPs shown in Fig. 4. Two intermolecular H-bonds formed among water molecules can be found in TA-3W-1 and TA-3W-2, respectively. For either TA-3W-3 or TA-3W-4, only one intermolecular H-bond is formed between two of three water molecules, and no hydrogen bonding interaction exists between the third water and the other two water molecules. The strongest intermolecular H-bond in TA-3W-2 is the $\text{O1H1}^{\text{TA}}\cdots\text{O}^{\text{C}}$ H-bond formed between TA and one water molecule, while one of the H-bonds formed between two water molecules is strongest in the other three complexes. Similar to TA-2W-2, because no intramolecular H-bond is formed in TA-3W-2, the cooperative effect occurring in TA-3W-2 strengthens the intermolecular $\text{O1H1}^{\text{TA}}\cdots\text{O}^{\text{C}}$ H-bond, which also differs from those occurring in other complexes, in which the cooperative effect enhances the strength of the intramolecular H-bonds.

Many-body interaction analysis

A quantitative account of the roles of hydrogen bonding interactions on $\text{TA}-(\text{H}_2\text{O})_n$ complexes was obtained through the energy decomposition scheme mentioned in the methods section on **Many-body interaction analysis**. The relaxation energies ($E_{\text{relaxation}}$), n -body interaction energies ($E(2\text{B})$, $E(3\text{B})$ and $E(4\text{B})$) and binding energies (E_{bind}) are listed in Table 4. Note that the $E_{\text{relaxation}}$ values were determined not in their own basis sets of monomers but in the basis set of the whole complex, while the energies of free monomers (E_{TA} and E_{W}) were determined in their own basis set rather than in the basis set of the whole complex. It is this difference of basis sets that results in the negative relaxation energies of some complexes as shown in Table 4. Therefore, the negative $E_{\text{relaxation}}$ in Table 4 does not mean that the TA/water moieties in complexes are more stable than their free monomers. In general, slight deformation results in small positive, or even negative, $E_{\text{relaxation}}$, while the serious deformation occurring in the complex is responsible for the larger and positive $E_{\text{relaxation}}$ of some complexes, including TA-1W-1, TA-1W-4, TA-2W-2, TA-3W-1 and TA-3W-2.

As shown in Table 4, the major contributions to the binding energies (E_{bind}) of $\text{TA}-(\text{H}_2\text{O})$ complexes come from

Table 4 Many-body interaction energies (two, three and four body) for $\text{TA}-(\text{H}_2\text{O})_n$ ($n=1-3$) complexes using $\omega\text{B97XD}/6-311++\text{G}(\text{d,p})$ level. All energies are in $\text{kcal}\cdot\text{mol}^{-1}$ and BSSE corrected

Complex	$E_{\text{relaxation}}$	$E(2\text{B})$	$E(3\text{B})$	$E(4\text{B})$	E_{bind}
TA-1W-1	0.26	-12.54			-12.28
TA-1W-2	-0.31	-9.80			-10.11
TA-1W-3	-0.54	-8.31			-8.85
TA-1W-4	6.30	-13.74			-7.43
TA-2W-1	-0.65	-20.27	-3.77		-24.69
TA-2W-2	3.63	-22.43	-3.44		-22.23
TA-2W-3	-0.11	-20.81	0.27		-20.65
TA-2W-4	-0.24	-20.41	0.23		-20.43
TA-3W-1	8.84	-42.12	-8.68	-0.22	-42.18
TA-3W-2	2.76	-30.40	-6.04	-0.40	-34.08
TA-3W-3	-0.56	-28.69	-3.62	0.01	-32.87
TA-3W-4	-0.03	-29.74	-3.54	0.03	-33.28

the two-body energies. TA-1W-1 is the most stable $\text{TA}-(\text{H}_2\text{O})$ complex since it involves the lowest E_{bind} of $-12.28 \text{ kcal}\cdot\text{mol}^{-1}$. However, although the $E(2\text{B})$ of $-13.74 \text{ kcal}\cdot\text{mol}^{-1}$ for TA-1W-4 is the lowest among the four $\text{TA}-(\text{H}_2\text{O})$ complexes, the largest positive $E_{\text{relaxation}}$ of $6.30 \text{ kcal}\cdot\text{mol}^{-1}$ counteracts the $E(2\text{B})$ to a great extent and results in the largest E_{bind} of $-7.43 \text{ kcal}\cdot\text{mol}^{-1}$. The $E_{\text{relaxation}}$ is also part of binding energy and is responsible for the structural deformation of complexes compared with free monomers. As shown in Table 1, the intramolecular H-bond in TA-1W-4 is the $\text{NH3}^{\text{TA}}\cdots\text{O1}^{\text{TA}}$ H-bond instead of the $\text{O1H1}^{\text{TA}}\cdots\text{N}^{\text{TA}}$ H-bond, which is different from that of the free TA molecule, while the intramolecular $\text{O1H1}^{\text{TA}}\cdots\text{N}^{\text{TA}}$ H-bonds are retained in other three $\text{TA}-(\text{H}_2\text{O})$ complexes, so that the structural deformation of TA-1W-4 is more serious than the other $\text{TA}-(\text{H}_2\text{O})$ complexes, which is consistent with the order of $E_{\text{relaxation}}$. Therefore, TA inclines to strengthen the intramolecular $\text{O1H1}^{\text{TA}}\cdots\text{N}^{\text{TA}}$ H-bond when it meets with the first water molecule.

For $\text{TA}-(\text{H}_2\text{O})_2$ complexes, the major contributions to the binding energies still come from the two-body energies $E(2\text{B})$. The three-body energies $E(3\text{B})$ favor the binding energies in TA-2W-1 and TA-2W-2, while $E(3\text{B})$ of TA-2W-3 and TA-2W-4 do not favor the binding energies. Moreover, the absolute values of $E(3\text{B})$ of both TA-2W-1 and TA-2W-2 are larger than those of the other two $\text{TA}-(\text{H}_2\text{O})_2$ complexes, which can be attributed partially to the greater number of hydrogen bonding interactions among water molecules existing in both TA-2W-1 and TA-2W-2. The largest $E_{\text{relaxation}}$ of $3.63 \text{ kcal}\cdot\text{mol}^{-1}$ of TA-2W-2 indicates the occurrence of serious structural deformation, as can be seen from Table 2 since the intramolecular $\text{O1H1}^{\text{TA}}\cdots\text{N}^{\text{TA}}$ H-bond is destroyed. The intramolecular $\text{O1H1}^{\text{TA}}\cdots\text{N}^{\text{TA}}$ H-bonds in other $\text{TA}-(\text{H}_2\text{O})_2$ complexes remain intact and are further

strengthened compared with those in TA-(H₂O) complexes. TA-2W-1 is the most stable TA-(H₂O)₂ complex since it involves the lowest binding energy (E_{bind}) of $-24.69 \text{ kcal}\cdot\text{mol}^{-1}$. The intramolecular O1H1^{TA}...N^{TA} H-bond in TA-2W-1 is also the strongest one among the TA-(H₂O)₂ complexes since the proton is closer to the N atom than those in other TA-(H₂O)₂ as well as TA-(H₂O) complexes, which indicates the very obvious trend of transformation of the zwitterionic complex from neutral complexes. Therefore, the conformational preferences of TA-(H₂O)₂ complexes are those that keep the intramolecular O1H1^{TA}...N^{TA} H-bond untouched and prefer to form further zwitterionic complexes.

Of the TA-(H₂O)₃ complexes, TA-3W-1 is the most stable since it involves the lowest E_{bind} of $-42.18 \text{ kcal}\cdot\text{mol}^{-1}$, and it is also the complex involving most serious deformation with $E_{\text{relaxation}}$ of $8.84 \text{ kcal}\cdot\text{mol}^{-1}$ since the intramolecular H-bond in TA-3W-1 is the NH1^{TA}...O1^{TA} rather than O1H1^{TA}...N^{TA} H-bond, which changes the complex from neutral to zwitterionic. Moreover, intramolecular O1H1^{TA}...N^{TA} H-bonds still exist in both TA-3W-3 and TA-3W-4 and are further strengthened compared to those in TA-(H₂O)₂ complexes, and the trend of transformation from neutral to zwitterionic forms is also found here. The cleavage of the intramolecular O1H1^{TA}...N^{TA} H-bond in TA-3W-2 results in a positive $E_{\text{relaxation}}$ of $2.76 \text{ kcal}\cdot\text{mol}^{-1}$, while slight structural deformations of other two TA-(H₂O)₃ complexes are confirmed by the small values of $E_{\text{relaxation}}$. Similar to TA-(H₂O)₂ complexes, which are partial due to the greater number of hydrogen bonding interactions among water molecules, the three-body energies of both TA-3W-1 and TA-3W-2 favor the binding energies, while fewer hydrogen bonding interactions among water molecules in both TA-3W-3 and TA-3W-4 result in little contribution of $E(3B)$ to their binding energies. The four-body energies of all TA-(H₂O)₃ complexes are very small and their contribution to the binding energies can be ignored.

Conclusions

The microsolvation of TA with water molecules was studied at the $\omega\text{B97XD/6-311++G(d,p)}$ level. The geometries and energies of TA-(H₂O)_{*n*} (*n*=1–3) complexes have been systematically investigated. QTAIM analyses have also been applied to understand the nature of the hydrogen bonding interactions in complexes. The results showed that intramolecular O1H1^{TA}...N^{TA} H-bonds remained untouched in most TA-(H₂O)_{*n*} (*n*=1–3) complexes, and are strengthened via the cooperative effects among multiple H-bonds from *n* = 1 to 3. Therefore, when TA interacts with more water molecules, a trend of transformation from the intramolecular O1H1^{TA}...N^{TA} H-bond to the NH1^{TA}...O1^{TA}

bond exists. Because of the formation of the intramolecular NH1^{TA}...O1^{TA} H-bond, the most stable TA-(H₂O)₃ complex, TA-3W-1, becomes a zwitterionic complex rather than neutral complex. Many-body interaction analysis showed that the major contributions to the binding energies of these complexes come from the two-body energies, while the relaxation energies and the three-body energies make a significant contribution to the binding energies for some of the complexes, whereas four-body energies are too small to be significant. To summarize, hydrogen bonding interactions between TA and water molecules play a key role in the microsolvations and conformations of TA, and result in the transformation from neutral to zwitterionic character. Of course, the present paper concentrates mainly on intermolecular interactions and the transformation from the neutral complex to the zwitterionic type. Therefore, in order to reduce the number of complexes, complexes forming hydrogen bonds among water molecules were omitted. This research may be helpful in further experimental and theoretical studies on the interactions of TA with different solvent molecules.

Acknowledgments This work is supported by Tianjin Science and Technology Development Fund Projects in Colleges and Universities (No. 20080504).

References

1. Sturman JA (1993) Taurine in development. *Physiol Rev* 73:119–147
2. Huxtable RJ (1992) Physiological actions of taurine. *Physiol Rev* 72:101–163
3. del Olmo N, Bustamante J, del Rio RM, Solis JM (2000) Taurine activates GABA(A) but not GABA(B) receptors in rat hippocampal CA1 area. *Brain Res* 864:298–307
4. Jia F, Yue M, Chandra D, Keramidias A, Goldstein PA, Homanics GE, Harrison NL (2008) Taurine is a potent activator of extrasynaptic GABA(A) receptors in the thalamus. *J Neurosci* 28:106–115
5. Beukes JA, Mo F, van Beek W (2007) X-Ray induced radiation damage in taurine: a combined X-ray diffraction and Raman study. *Phys Chem Chem Phys* 9:4709–4720
6. Dewar MJS, Zebisch EG, Healy EF, Stewart JJP (1985) Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J Am Chem Soc* 107:3902–3909
7. Cortijo V, Sanz ME, Lopez JC, Alonso JL (2009) Conformational study of taurine in the gas phase. *J Phys Chem A* 113:14681–14683
8. Hibbs DE, Austin-Woods CJ, Platts JA, Overgaard J, Turner P (2003) Experimental and theoretical charge density study of the neurotransmitter taurine. *Chem Eur J* 9:1075–1084
9. Waller MP, Howard ST, Platts JA, Piltz RO, Willock DJ, Hibbs DE (2006) Novel properties from experimental charge densities: an application to the zwitterionic neurotransmitter taurine. *Chem Eur J* 12:7603–7614
10. Huang ZG, Dai YM, Yu L (2010) Density functional theory and topological analysis on the hydrogen bonding interactions in N-protonated adrenaline-DMSO complexes. *Struct Chem* 21:863–872

11. Huang ZG, Yu L, Dai YM (2010) Combined DFT with NBO and QTAIM studies on the hydrogen bonds in $(\text{CH}_3\text{OH})_n$ ($n=2-8$) clusters. *Struct Chem* 21:565–572
12. Rappe AK, Bernstein ER (2000) Ab initio calculation of nonbonded interactions: are we there yet? *J Phys Chem A* 104:6117–6128
13. Grimme S (2004) Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J Comput Chem* 25:1463–1473
14. Schwabe T, Grimme S (2007) Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability. *Phys Chem Chem Phys* 9:3397–3406
15. Zhao Y, Truhlar DG (2006) Comparative DFT study of van der Waals complexes: rare-gas dimers, alkaline-earth dimers, zinc dimer, and zinc-rare-gas dimers. *J Phys Chem A* 110:5121–5129
16. Zhao Y, Truhlar DG (2008) The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Acc* 120:215–241
17. Chai JD, Head-Gordon M (2008) Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys Chem Chem Phys* 10:6615–6620
18. Johnson ER, Mackie ID, DiLabio GA (2009) Dispersion interactions in density-functional theory. *J Phys Org Chem* 22:1127–1135
19. Riley KE, Pitonak M, Jurecka P, Hobza P (2010) Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chem Rev* 110:5023–5063
20. Mandal A, Prakash M, Kumar RM, Parthasarathi R, Subramanian V (2010) Ab initio and DFT studies on methanol-water clusters. *J Phys Chem A* 114:2250–2258
21. Rai AK, Fei WX, Lu ZW, Lin ZJ (2009) Effects of microsolvation and aqueous solvation on the tautomers of histidine: a computational study on energy, structure and IR spectrum. *Theor Chem Acc* 124:37–47
22. Rao L, Ke HW, Fu G, Xu X, Yan YJ (2009) Performance of several density functional theory methods on describing hydrogen-bond interactions. *J Chem Theor Comput* 5:86–96
23. Riley KE, Pitonak M, Cerny J, Hobza P (2010) On the structure and geometry of biomolecular binding motifs (hydrogen-bonding, stacking, $\text{X-H}\cdots\pi$): WFT and DFT calculations. *J Chem Theor Comput* 6:66–80
24. Huang ZG, Yu L, Dai YM, Wang HK (2010) Hydrogen bonding interactions in cysteine-urea complexes: theoretical studies of structures, properties and topologies. *J Mol Struct THEOCHEM* 960:98–105
25. Huang ZG, Dai YM, Wang HK, Yu L (2011) Microsolvation of aminoethanol: A DFT combined with QTAIM study. *J Mol Model*. doi:10.1007/s00894-011-0973-z
26. Huang ZG, Dai YM, Yu L, Wang HK (2011) Hydrogen bonding interactions in noradrenaline- DMSO complexes: DFT and QTAIM studies of structure, properties and topology. *J Mol Model*. doi:10.1007/s00894-011-0956-0
27. Plumley JA, Dannenberg JJ (2011) A comparison of the behavior of functional/basis set combinations for hydrogen-bonding in the water dimer with emphasis on basis set superposition error. *J Comput Chem* 32. doi:10.1002/jcc.21729
28. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian09. Gaussian Inc, Wallingford
29. McLean AD, Chandler GS (1980) Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, $Z=11-18$. *J Chem Phys* 72:5639–5648
30. Krishnan R, Binkley JS, Seeger R, Pople JA (1980) Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J Chem Phys* 72:650–654
31. Boys SF, Bernardi F (1970) The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol Phys* 19:553–566
32. Biegler-König F (2000) AIM2000. 1.0 edn. University of Applied Sciences, Bielefeld, Germany
33. Huang ZG, Yu L, Dai YM (2010) Density functional theory and topological analysis on the hydrogen bonds in cysteine-propanoic acid complexes. *Struct Chem* 21:855–862
34. Nozad AG, Meftah S, Ghasemi MH, Kiyani RA, Aghazadeh M (2009) Investigation of intermolecular hydrogen bond interactions in crystalline L-cysteine by DFT calculations of the oxygen-17, nitrogen-14, and hydrogen-2 EFG tensors and AIM analysis. *Biophys Chem* 141:49–58
35. Zhou HW, Lai WP, Zhang ZQ, Li WK, Cheung HY (2009) Computational study on the molecular inclusion of andrographolide by cyclodextrin. *J Comput Aided Mol Des* 23:153–162
36. Koch U, Popelier PLA (1995) Characterization of C-H-O hydrogen bonds on the basis of the charge density. *J Phys Chem* 99:9747–9754
37. Maheshwary S, Patel N, Sathyamurthy N, Kulkarni AD, Gadre SR (2001) Structure and stability of water clusters $(\text{H}_2\text{O})_n$, $n = 8-20$: an ab initio investigation. *J Phys Chem A* 105:10525–10537
38. Xantheas SS (2000) Cooperativity and hydrogen bonding network in water clusters. *Chem Phys* 258:225–231
39. Chaudhari A, Lee SL (2004) A computational study of microsolvation effect on ethylene glycol by density functional method. *J Chem Phys* 120:7464–7469
40. Chaudhari A, Sahu PK, Lee SL (2004) Many-body interaction in glycine-(water)₃ complex using density functional theory method. *J Chem Phys* 120:170–174
41. Kulkarni AD, Ganesh V, Gadre SR (2004) Many-body interaction analysis: algorithm development and application to large molecular clusters. *J Chem Phys* 121:5043–5050
42. Hannachi Y, Silvi B, Bouteiller Y (1992) Ab initio study of the structure, cooperativity, and vibrational properties of the $\text{H}_2\text{O}:(\text{HF})_2$ hydrogen bonded complex. *J Chem Phys* 97:1911–1918
43. Mo O, Yanez M, Elguero J (1992) Cooperative (nonpairwise) effects in water trimers: an ab initio molecular orbital study. *J Chem Phys* 97:6628–6638
44. Frank HS, Wen WY (1957) Ion-solvent interaction. Structural aspects of ion-solvent interaction in aqueous solutions: a suggested picture of water structure. *Discuss Faraday Soc* 24:133–140

Ab initio calculation of the geometries, stabilities, and electronic properties for the bimetallic Be_2Au_n ($n=1-9$) clusters: comparison with pure gold clusters

Ya-Ru Zhao · Xiao-Yu Kuang · Bao-Bing Zheng ·
Su-Juan Wang · Yan-Fang Li

Received: 22 November 2010 / Accepted: 18 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract Ab initio methods based on density functional theory at BP86 level were applied to the study of the geometrical structures, relative stabilities, and electronic properties of small bimetallic Be_2Au_n ($n=1-9$) clusters. The optimized geometries reveal that the most stable isomers have 3D structures at $n=3, 5, 7, 8,$ and 9 . Here, the relative stabilities were investigated in terms of the averaged atomic binding energies, fragmentation energies and second-order difference of energies. The results show that the planar Be_2Au_4 structure is the most stable structure for Be_2Au_n clusters. The HOMO–LUMO gap, vertical ionization potential, vertical electron affinity and chemical hardness exhibit a pronounced even–odd alternating phenomenon. In addition, charge transfer and natural electron configuration were analyzed and compared.

Keywords Be–Au cluster · Geometric configuration · Density functional method

Introduction

As with all simple small-sized clusters, bimetallic clusters—as a suitable model for computational quantum chemistry—

often have properties that differ markedly from both the atomic state and the bulk material [1–6]. Pure and doped gold clusters in particular have received much attention because of their particular physical and chemical properties, and because of their potential technological applications in solid state chemistry, materials science, optics, nanotechnology, catalysis, biology and medicine [7–21].

On the experimental side, a great number of studies have been performed on alloyed gold clusters. For instance, AgAu and CuAu dimers have been studied using resonant two-photon ionization spectroscopy [22]. Janssens et al. [23], using cationic photofragmentation mass spectrometry, found strongly enhanced stabilities for $[\text{Au}_5X]^+$ ($X = \text{V}, \text{Mn}, \text{Cr}, \text{Fe}, \text{Co}, \text{Zn}$) clusters. Negishi et al. [24] investigated the photoelectron spectroscopy (PES) of Ag_mAu_n^- ($m+n \leq 4$) clusters. Koyasu et al. [25] conducted a systematic investigation into the anion PES of small metal-doped bimetallic Au clusters, $[\text{Au}_nM]^-$ ($n=2-7$; $M = \text{Pd}, \text{Ni}, \text{Zn}, \text{Cu},$ and Mg), the results of which have shown that, except for Au_nPd^- clusters, the spectra of electron affinities exhibit even–odd alternation. Heiz et al. [26] investigated the gold-doped alkali-metal clusters Na_xAu and Cs_xAu , measuring the ionization potentials of these clusters and performing theoretical calculations at the relativistic density-functional level. Their results indicate that the electronic structures of Na_xAu clusters are jelliumlike, whereas Cs_xAu clusters show significant ionic character.

Theoretically, Heinebrodt et al. [27] investigated bimetallic Au_nX_m ($X = \text{Cu}, \text{Al}, \text{Y}, \text{In}, \text{Cs}$) clusters, and found electronic shell effects. By investigating the Au_nM ($n=1-7$, $M = \text{Ni}, \text{Pd}, \text{Pt}$) clusters, Yuan et al. [28] found that the presence of atoms as impurities can influence markedly the geometric and electronic properties of gold clusters. The structures and electronic properties of $\text{Au}_{19}X$ clusters ($X = \text{Li}, \text{Na}, \text{K}, \text{Ru}, \text{Cs}, \text{Cu},$ and Ag) have been studied by

Y.-R. Zhao · X.-Y. Kuang (✉) · S.-J. Wang · Y.-F. Li
Institute of Atomic and Molecular Physics, Sichuan University,
Chengdu 610065, China
e-mail: scu_kuang@163.com

X.-Y. Kuang
International Centre for Materials Physics,
Chinese Academy of Sciences,
Shenyang 110016, China

B.-B. Zheng
Department of Physics, Baoji University of Arts and Science,
Baoji 721007, China

Ghanty et al. [29]. Recently, Guo et al. studied the structures and stabilities of Au_nPd_2 and Au_nPt_2 ($n=1-4$) clusters [30, 31]; the results indicated that gold-doped atom interactions are strong enough to improve cluster stability, and the larger the Au_n cluster, smaller the distortion by two Pd-, Pt-atoms substitutions. For beryllium-gold clusters, the dissociation energy, vibration frequency and bond length of the BeAu dimer were reported by Balducci et al. [32] based on mass spectrometry and density functional theory (DFT). However, to the best of our knowledge, few systematic theoretical investigations into Be_2Au_n ($n=1-9$) clusters have been performed. For example, if two beryllium atoms are doped in gold clusters, do their structures and properties differ from those of bare gold clusters? In the present study, we report a systematic DFT investigation into small bimetallic Be_2Au_n ($n=1-9$) clusters, comparing these with pure gold clusters. We find some interesting tendencies in the geometric structures of Be_2Au_n ($n=1-9$) clusters. Furthermore, averaged atomic bonding energies and second-order energy difference show that the Be_2Au_4 isomer is the most stable structure among Be_2Au_n clusters. HOMO–LUMO gaps, vertical ionization potential, vertical electron affinity and chemical hardness display an even–odd alternation with cluster size. In addition, charge transfer and natural electron configuration were also analyzed and further compared. It is hoped that our theoretical study will be useful not only in deepening our understanding of the influence of local structure on material properties, but also in providing powerful guidelines for future experimental research.

Theoretical methods and computational details

All optimizations of Be_2Au_n ($n=1-9$) clusters were performed using the GAUSSIAN 03 program package [33] at BP86 level [34–36]. For the Au atom, full electron calculation is rather time-consuming, so it is better to introduce effective core potential including relativistic effects (RECP) to describe the inner-core electrons. Basis sets labeled GENECP are then combinations of LANL2TZ

(f) [37, 38] and 6-311+G (d) [39] basis sets, which are used for Au and Be atoms, respectively. In the search for the lowest energy structure, lots of possible initial structures, including one-, two- and three-dimensional configurations, were considered in geometry optimization, and all the clusters were relaxed fully without any symmetry constraints. Regarding spin polarization, different spin multiplicities were also taken into account for each initial configuration. Furthermore, vibration frequency calculations were performed to guarantee the optimized structures corresponding to a local minimum and provide zero-point energy (ZPE). In this way, a large number of optimized isomers for Be_2Au_n ($n=1-9$) clusters were obtained, but here we report only a few energetically low-lying isomers for each size.

To test the reliability of our calculations, the bond lengths, vibration frequencies, and dissociation energies of BeAu and Au_2 dimers were calculated (listed in Table 1). It can be seen that the results based on the BP86 method agree more with experimental values [32, 40, 41] than the others. This indicates the suitability of the chosen computational method to describe small Be_2Au_n ($n=1-9$) clusters. In addition, the vertical ionization potential ($IP_v = E_{\text{cation at optimized neutral geometry}} - E_{\text{optimized neutral}}$), vertical electron affinity ($EA_v = E_{\text{optimized neutral}} - E_{\text{anion at optimized neutral geometry}}$), highest occupied–lowest unoccupied molecular orbital (HOMO–LUMO) energy gap, and natural population analysis (NPA) of the stable configurations were also performed based at the BP86 level.

Results and discussion

Bare gold clusters Au_n ($n=3-11$)

In order to compare the effects of doped impurity atoms on gold clusters, we first performed some optimizations and discussions on Au_n ($n=3-11$) clusters. Taking many possible initial structures into account, only the most stable geometric isomers of gold clusters for each size are shown in Fig. 1. In addition, the averaged atomic binding energy,

Table 1 Calculated and experimental bond lengths (r), dissociation energies (D_e), and frequencies (ω_e) for the AuBe and Au_2 clusters

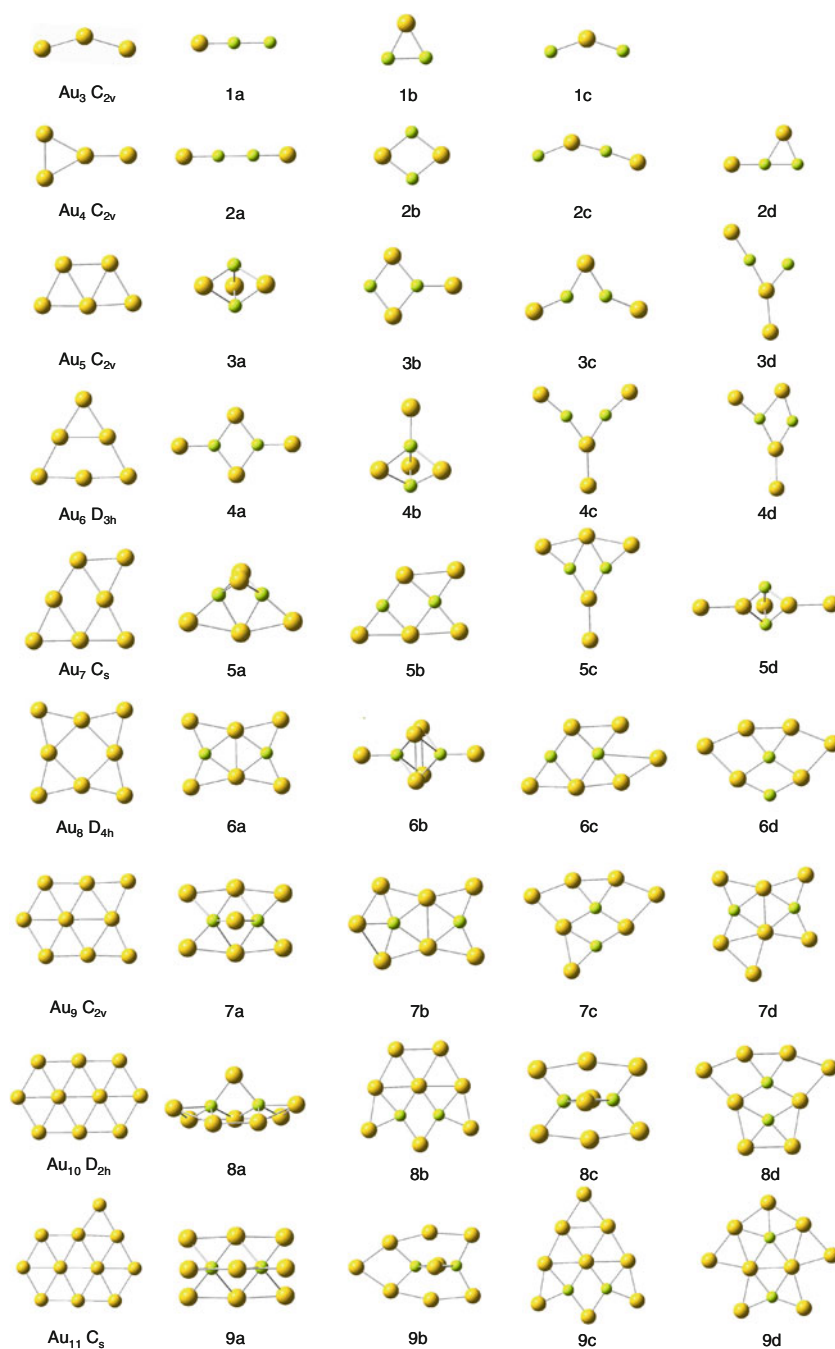
Methods	AuBe			Au ₂		
	r (Å)	D_e (eV)	ω_e (cm ⁻¹)	r (Å)	D_e (eV)	ω_e (cm ⁻¹)
PW91	2.05	2.72	608.60	2.52	2.30	176.60
PW91P86	2.05	2.77	609.66	2.52	2.43	178.00
B3LYP	2.06	2.30	592.92	2.55	1.98	169.47
BP86	2.06	2.65	605.54	2.52	2.28	176.21
Experimental	2.06 ^a	2.9 ^a 2.425 ^b	607.68 ^a	2.47 ^c	2.29 ^c	191.0 ^c

^a[37]

^b[32]

^c[38]

Fig. 1 Lowest energy structures of Au_{n+2} and Be_2Au_n ($n=1-9$) clusters, and few low-lying isomers for doped clusters. Yellow and green balls represent Au and Be atoms, respectively



second-order difference of energy, HOMO–LUMO gap, IP_v , and EA_v and hardness of small size gold clusters were studied and compared with the available experimental values in the following.

Bimetallic beryllium-gold clusters Be_2Au_n ($n=1-9$)

For Be_2Au_n ($n=1-9$) clusters, the ground state isomer and few low-lying structures for each size are displayed in Fig. 1. According to the total energies from low to high,

these isomers are designated by na , nb , nc , and nd . (“ n ” is the number of Au atoms in the Be_2Au_n clusters) Meanwhile, Table 2 lists the electronic states, symmetries, relative energies, HOMO and LUMO energies, as well as vibration frequencies of Be_2Au_n clusters.

Possible Be_2Au geometries such as C_{2v} and $D_{\infty h}$ isomers are optimized as stable structures. According to the calculated results, the lowest energy isomer is a linear structure with $C_{\infty v}$ symmetry, in which Be–Be and Be–Au bond lengths are 2.15 Å and 2.06 Å, respectively. However,

Table 2 Electronic states, symmetries, relative energies (ΔE), HOMO energies, LUMO energies, and vibration frequencies of Be_2Au_n ($n=1-9$) clusters

Isomer	State	Symmetry	ΔE (eV)	HOMO (Hartree)	LUMO (Hartree)	Frequency (cm^{-1})
1a	$^2\Sigma$	$C_{\infty v}$	0.00	-0.15817	-0.11692	144.1, 328.1, 775.2
1b	2B_2	C_{2v}	0.05	-0.18269	-0.13704	356.1, 385, 631.2
1c	2A_1	C_{2v}	0.28	-0.16869	-0.12217	67.2, 399.9, 509.9
2a	$^1\Sigma_g$	$D_{\infty h}$	0.00	-0.21835	-0.11180	83.4, 93.6, 272, 608.3, 892.2
2b	1A_g	D_{2h}	0.08	-0.17398	-0.11810	93.2, 171.6, 364.8, 484.8, 501.6
2c	$^1A'$	C_s	0.71	-0.19822	-0.12723	47.5, 118.1, 180.8, 196.3, 409.7
2d	$^1A'$	C_s	0.75	-0.19305	-0.14471	30.5, 163.3, 211.9, 314.8, 491.9
3a	–	D_{3h}	0.00	-0.17413	-0.14560	53.2, 54.1, 115.6, 338, 389.4
3b	2A_1	C_{2v}	0.36	-0.16461	-0.12919	22.1, 48.6, 66.6, 107.8, 206.4
3c	2A_1	C_{2v}	0.59	-0.18906	-0.15237	24.2, 53.5, 100.3, 103.2, 211.6
3d	–	C_s	1.92	-0.19699	-0.15386	23.5, 47.9, 100.5, 104.7, 174.5
4a	1A_g	D_{2h}	0.00	-0.21840	-0.12145	12.1, 13, 27.2, 53.8, 96.9
4b	1A_1	C_{3v}	0.13	-0.22076	-0.13656	18.7, 52, 84, 113.6, 306.7
4c	$^1A'$	C_s	1.29	-0.20362	-0.14773	13.4, 21.5, 37.6, 92, 101.7
4d	1A_1	C_{2v}	1.33	-0.17742	-0.11759	16.3, 17, 26.1, 99.9, 110.5
5a	2A_1	C_{2v}	0.00	-0.17785	-0.16397	25.7, 27.2, 38.2, 43.5, 68.4
5b	$^2A'$	C_s	0.28	-0.18059	-0.16771	9.3, 26.6, 33.7, 57.7, 62.8
5c	2A_1	C_{2v}	1.04	-0.19642	-0.18386	14.6, 20.6, 25.7, 55.1, 84.5
5d	2B_1	C_{2v}	1.96	-0.19377	-0.16510	18.2, 20.5, 23.7, 46.6, 67.3
6a	1A_g	D_{2h}	0.00	-0.20286	-0.14183	13.3, 24.2, 45.9, 47.3, 48.2
6b	$^1A_{1g}$	D_{4h}	0.49	-0.21334	-0.16142	17.2, 20.9, 34.4, 66.9, 68.8
6c	–	C_s	0.74	-0.19103	-0.15284	15.2, 19.4, 32.5, 47.4, 49.8
6d	1A_1	C_{2v}	0.89	-0.21947	-0.14661	19.6, 25.7, 48.6, 50, 61.1
7a	2A_1	C_{2v}	0.00	-0.18911	-0.17986	16, 28.8, 32.1, 36.6, 45.9
7b	$^2A'$	C_s	0.42	-0.17883	-0.16932	10.3, 12.5, 26.9, 38.9, 41.3
7c	$^2A'$	C_s	0.65	-0.19660	-0.18545	15.4, 22, 32.7, 34.9, 48.3
7d	$^2A'$	C_s	0.76	-0.19474	-0.18222	14.2, 16.7, 25.2, 41.1, 50.5
8a	$^1A'$	C_s	0.00	-0.21313	-0.14443	9.8, 21.1, 28.9, 33.8, 34.1
8b	1A_1	C_{2v}	0.31	-0.21055	-0.14914	15.8, 18.2, 33, 37.2, 39.3
8c	1A_g	D_{2h}	0.33	-0.19994	-0.15713	11.2, 23.1, 26.5, 32.8, 35.5
8d	1A_1	C_{2v}	0.34	-0.21396	-0.15974	15.6, 18.3, 25.9, 28.9, 44.5
9a	–	D_{3h}	0.00	-0.20037	-0.19276	13.2, 18.6, 20.1, 34.8, 36.6
9b	2A_1	C_{2v}	0.12	-0.18940	-0.18277	11.2, 16.9, 25.6, 29.9, 31.8
9c	2A_1	C_{2v}	0.30	-0.18562	-0.17767	12.6, 18.5, 22.3, 28.1, 33.4
9d	2B_2	C_{2v}	0.34	-0.18026	-0.17319	17.8, 24.6, 28.2, 35.1, 40.2

the substituted structure (1c) of ground-state Au_3 clusters is higher in energy than that of the linear structure by 0.28 eV. For Be_2Au_2 clusters, the linear structure is also found to be the most stable structure. This structure, with $D_{\infty h}$ symmetry and $^1\Sigma_g$ electronic state, is obtained when an Au atom is added to the 1a isomer. The Be–Be bond length in the 2a isomer shortens to 2.08 Å due to the effects of the doped Au atom. Our calculations reveal that the Y-type structure is the ground state isomer of Au_4 clusters. However, the Y-type structure of Be_2Au_2 cluster is 3.45 eV higher in energy than that of 2a structure. Stating at $n=3$, the lowest energy

structure of the Be_2Au_n ($n=1-9$) cluster shows the appearance of 3D geometries. The trigonal bipyramid (3a), which has D_{3h} symmetry, proved the most stable isomer of Be_2Au_3 clusters. After the Au atom is capped on the scoop-shaped structure 3b, the most stable structure 4a, which has a planer structure, is generated. Unfortunately, the lowest energy 3D isomer (4b) is 0.13 eV higher in total energy than the most stable isomer 4a. Among the Be_2Au_5 clusters, the 3D isomers 5a and 5d with C_{2v} symmetry can be viewed as two isomers derived from 3a, but in which the fourth and fifth Au atoms cap on different atoms of the

trigonal bipyramid 3a structure. Therefore, different relative energy (1.96 eV), and electronic states (2A_1 , 2B_1) exist between them. As shown in Fig. 1, the ground state Au_8 cluster is a planar structure with D_{4h} symmetry. Here, the substituted isomer (6a) is the most stable structure for Be_2Au_6 clusters. Due to the effect of the Be atoms, the symmetry of 6a is lowered, i.e., distorted from D_{4h} to D_{2h} . The results of total energy calculations reveal that the 3D

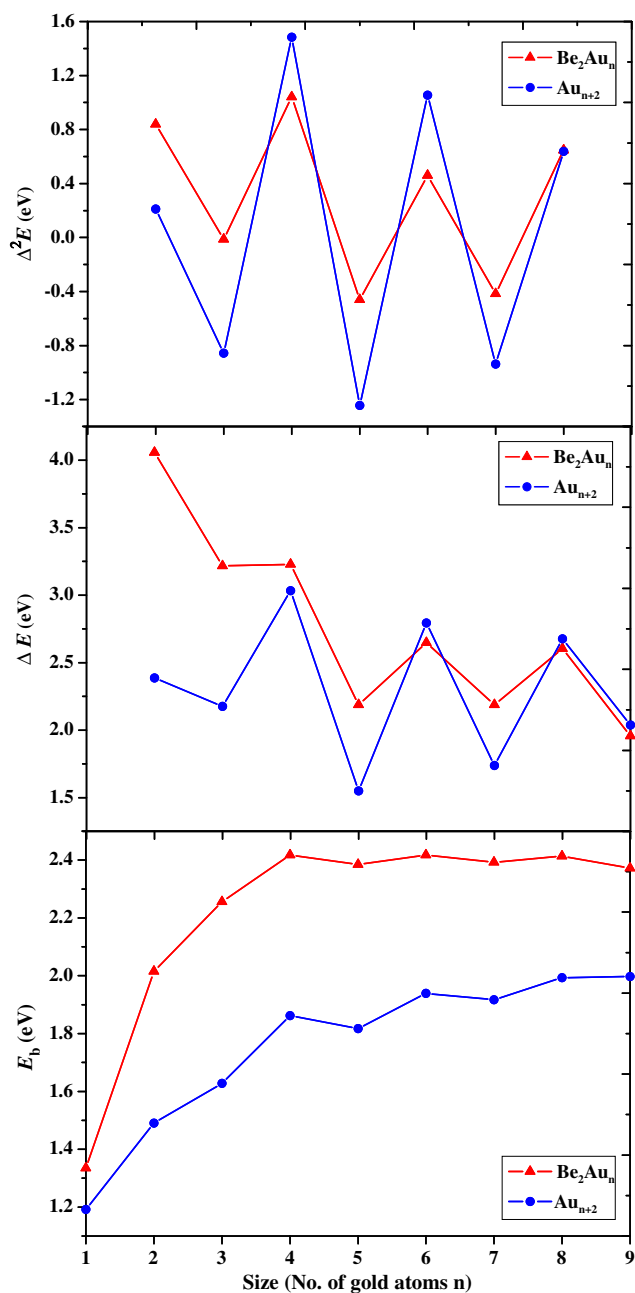


Fig. 2 Size dependence of the average atomic binding energies $E_b(n)$, fragmentation energies $\Delta E(n)$, and the second-order difference of energies $\Delta_2 E(n)$ for the lowest energy structures of Au_{n+2} and Be_2Au_n ($n=1-9$) clusters

isomer (6b) is less stable than 6a by 0.49 eV. On the basis of the most stable structure of Be_2Au_6 clusters, three derived structures (7a, 7b, and 7d) are obtained when Au atoms are capped on different sites of the 6a structure. Among them, the 3D isomer (7a) is more stable than the other planar structures. Similarly, two low-lying 3D Be_2Au_8 structures (8a and 8c) are optimized and guided by the ground state Be_2Au_7 clusters. The most stable isomer is obtained by bi-capping Au atoms onto two Au atoms of 7b isomer while the 8c isomer is formed by bi-capping the Au atom onto two Be atoms. When the number of Be_2Au_n clusters is increased to 9, we still find the 3D isomer (9a) to be the lowest-energy structure. Interestingly, after two Au atoms are capped on the 7a isomer, the most stable structure of Be_2Au_9 clusters is obtained and the symmetry changes from C_{2v} to D_{3h} .

From the above discussion, it can be noted that the lowest energy structures of Be_2Au_n clusters for $n=3, 5, 7, 8, 9$ favor a 3D structure. Although Be_2Au_2 and Be_2Au_4 have planar structures, their structures are not similar to those of pure gold clusters. This indicates that the two doped Be atoms dramatically affect the geometries of the ground-state of Au_n clusters. In addition, one-Au-atom-capped Be_2Au_{n-1} structures for different sized Be_2Au_n ($n=1-9$) clusters are the most dominant growth pattern.

Relative stabilities

In order to predict the relative stabilities of Be_2Au_n clusters, the averaged atomic binding energies $E_b(n)$, fragmentation energies $\Delta E(n)$ and second-order difference of energies $\Delta_2 E(n)$ for different-sized Be_2Au_n and corresponding Au_n clusters were calculated. For Be_2Au_n

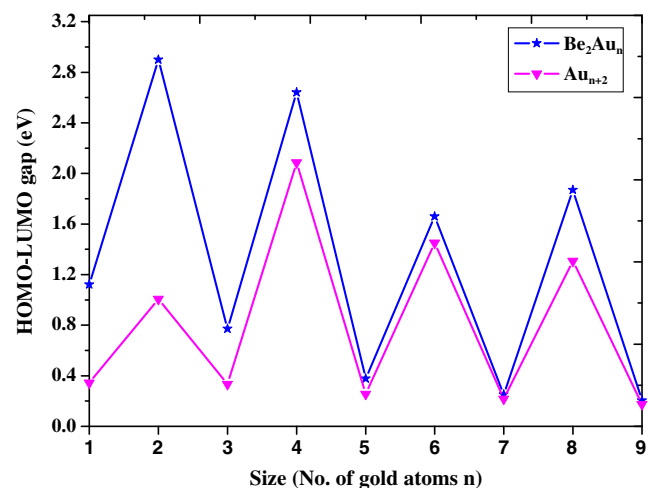


Fig. 3 Size dependence of the HOMO–LUMO gaps for the lowest energy structure of Au_{n+2} and Be_2Au_n ($n=1-9$) clusters

Table 3 Natural population analysis (NPA) of charges of the lowest energy Be_2Au_n ($n=1-9$) clusters

Isomers	Be-1	Be-2	Au-1	Au-2	Au-3	Au-4	Au-5	Au-6	Au-7	Au-8	Au-9
Be_2Au	0.189	0.076	-0.265								
Be_2Au_2	0.260	0.260	-0.260	-0.260							
Be_2Au_3	-0.063	-0.063	0.044	0.041	0.041						
Be_2Au_4	0.004	0.004	-0.090	-0.090	0.086	0.086					
Be_2Au_5	-0.327	-0.327	0.097	0.163	0.163	0.116	0.116				
Be_2Au_6	-0.160	-0.160	0.074	0.074	0.074	0.074	0.013	0.013			
Be_2Au_7	-0.636	-0.636	0.145	0.218	0.145	0.191	0.191	0.191	0.191		
Be_2Au_8	-0.650	-0.650	0.209	0.160	0.160	0.133	0.209	0.192	0.133	0.104	
Be_2Au_9	-1.034	-1.034	0.236	0.237	0.239	0.237	0.236	0.217	0.215	0.210	0.239

clusters, $E_b(n)$, $\Delta E(n)$, and $\Delta_2 E(n)$ were defined using the following formula:

$$E_b(n) = [2E(\text{Be}) + nE(\text{Au}) - E(\text{Be}_2\text{Au}_n)]/n + 2, \quad (1)$$

$$\Delta E(n) = E(\text{Be}_2\text{Au}_{n-1}) + E(\text{Au}) - E(\text{Be}_2\text{Au}_n), \quad (2)$$

$$\Delta_2 E(n) = E(\text{Be}_2\text{Au}_{n-1}) + E(\text{Be}_2\text{Au}_{n+1}) - 2E(\text{Be}_2\text{Au}_n), \quad (3)$$

where $E(\text{Be}_2\text{Au}_{n-1})$, $E(\text{Au})$, $E(\text{Be})$, $E(\text{Be}_2\text{Au}_n)$, and $E(\text{Be}_2\text{Au}_{n+1})$ denote the total energy of the $\text{Be}_2\text{Au}_{n-1}$, Au, Be, Be_2Au_n , and $\text{Be}_2\text{Au}_{n+1}$ clusters, respectively.

For Au_n clusters, $E_b(n)$, $\Delta E(n)$, and $\Delta_2 E(n)$ are defined by the following formula:

$$E_b(n) = [nE(\text{Au}) - E(\text{Au}_n)]/n, \quad (4)$$

$$\Delta E(n) = E(\text{Au}_{n-1}) + E(\text{Au}) - E(\text{Au}_n), \quad (5)$$

$$\Delta_2 E(n) = E(\text{Au}_{n-1}) + E(\text{Au}_{n+1}) - 2E(\text{Au}_n), \quad (6)$$

where $E(\text{Au}_{n-1})$, $E(\text{Au})$, $E(\text{Au}_n)$, and $E(\text{Au}_{n+1})$ denote the total energy of the Au_{n-1} , Au, Au_n , and Au_{n+1} clusters, respectively. The $E_b(n)$, $\Delta E(n)$, and $\Delta_2 E(n)$ values of the lowest energy Be_2Au_n and Au_{n+2} ($n=1-9$) clusters against the corresponding number of the Au atoms are plotted in Fig. 2. The features of the size evolution are easily seen, as are the peaks of the curves corresponding to those clusters having enhanced local stabilities. For Au_{n+2} clusters, the averaged atomic binding energy has an increasing tendency and $\Delta E(n)$, $\Delta_2 E(n)$ show obvious even-odd oscillations with increasing cluster size, which agrees with the previous works [9]. For Be_2Au_n clusters, the average atomic binding energy increases dramatically to a maximum as the size increases from $n=1$ to $n=4$; then, it shows a slight even-odd alternation at $n>4$. Therefore, a visible peak occurs at $n=4$, indicating that the Be_2Au_4 isomer is relatively more stable than its neighboring clusters. In addition, $E_b(n)$ values of Be_2Au_n clusters are significantly higher than those of Au_{n+2} clusters, which hints that the impurity Be atoms can strongly enhance the stabilities of gold clusters of small size. The fragmentation energies show a decreasing trend with the increase in cluster size and an even-odd oscillation occurs when $n>4$. Meanwhile, the second-order

Table 4 Natural electronic configurations of Be atoms in the most stable Be_2Au_n systems. Where Be (1) and Be (2) correspond to the top (or left) Be and bottom (or right) Be atoms in Fig. 1

Isomer	Be (1)				Be (2)			
	2s	2p	3p	3d	2s	2p	3p	3d
Be_2Au	1.33	0.47	0	0	1.20	0.72	0.01	0
Be_2Au_2	1.05	0.68	0	0	1.05	0.68	0	0
Be_2Au_3	0.78	1.25	0.02	0.01	0.78	1.25	0.02	0.01
Be_2Au_4	0.77	1.20	0	0.01	0.77	1.20	0	0.01
Be_2Au_5	0.66	1.63	0.03	0.01	0.66	1.63	0.03	0.01
Be_2Au_6	0.73	1.40	0	0.01	0.73	1.40	0	0.01
Be_2Au_7	0.65	1.94	0.03	0.01	0.65	1.94	0.03	0.01
Be_2Au_8	0.65	1.96	0	0.02	0.65	1.96	0	0.02
Be_2Au_9	0.68	2.31	0.03	0.01	0.68	2.31	0.03	0.01

difference of energies exhibits odd–even alternations in the region of $n=2-8$. This means that clusters containing an even number of gold atoms, especially Be_2Au_4 clusters, have higher relative stability than their neighbors.

HOMO–LUMO gaps and natural population analysis

The HOMO–LUMO energy gap represents the ability of a molecule to participate in a chemical reaction to some degree. A large value of the HOMO–LUMO energy gap is related to enhanced chemical stability; conversely, a small one corresponds to high chemical activity. In a sense this parameter can provide an important criterion to reflect the chemical stability of clusters. For the low-lying configurations of Be_2Au_n ($n=1-9$) clusters, HOMO and LUMO energies at each cluster size are listed in Table 1. We can obtain $E_g=1.12$ eV (Be_2Au), 2.90 eV (Be_2Au_2), 0.77 eV (Be_2Au_3), 2.64 eV (Be_2Au_4), 0.38 eV (Be_2Au_5), 1.66 eV (Be_2Au_6), 0.25 eV (Be_2Au_7), 1.87 eV (Be_2Au_8), and 0.20 eV (Be_2Au_9) for the most stable Be_2Au_n clusters. Meanwhile, the HOMO–LUMO gaps for the most stable Be_2Au_n and Au_{n+2} clusters are plotted against cluster size in Fig. 3. As can be seen from the figure, the tendency of the HOMO–LUMO gaps for Be_2Au_n clusters is similar to that of Au_{n+2} clusters, exhibiting an odd–even oscillatory behavior. Namely, clusters with an even number of atoms have enhanced chemical stability due to their larger gaps compared with their neighbors, which is in agreement with the results obtained with $E_b(n)$ and $\Delta_2E(n)$. Besides, we find that the HOMO–LUMO gaps of Be_2Au_n clusters are higher than those of Au_{n+2} clusters. This means that the impurity Be atoms can enhance the chemical stability of gold clusters.

NPA of Be_2Au_n ($n=1-9$) clusters can provide reliable charge-transfer information. Therefore, NPA results for the lowest energy Be_2Au_n species are summarized in Table 3. From this table, one can see that beryllium atoms possess positive charges for $n=1, 2, 4$ and negative charges for $n=3, 5-9$, suggesting that the charges in corresponding $\text{Be}_2\text{Au}_{1, 2, 4}$ clusters transfer from Be atoms to the Au_n frames, and that, in $\text{Be}_2\text{Au}_{3, 5-9}$, clusters transfer from Au_n atoms to Be atoms. Moreover, we find that the charges are different for each Be atom in some Be_2Au clusters, whereas they are equal in others. This may be due to the fact that there are unequal number of Be–Au bonds in each Be_2Au cluster. In other words, the charge distribution is dependent on the symmetry of the cluster. In order to understand internal charge transfer, the natural electron configurations for Be atoms in the most stable Be_2Au_n systems were investigated (results tabulated in Table 4). It was shown that the $2s$ states loses electrons 0.67–1.35, while the $2p$ and $3p$ states gain 0.47–2.34 electrons. The contribution of the $3d$ states is nearly zero. Furthermore, it is interesting that the

Table 5 Vertical ionization potential (IP_v), vertical electron affinity (EA_v), and chemical hardness (η) of the most stable Au_{n+2} and Be_2Au_n ($n=1-9$) clusters

Cluster size	Be_2Au_n			Au_{n+2}			
	η	EA_v	IP_v	η	EA_v	IP_v	IP_v^{29}
$n=1$	6.14	0.99	7.13	5.15	3.54	8.69	7.50
$n=2$	6.39	1.49	7.88	5.64	2.88	8.52	8.60
$n=3$	4.94	2.00	6.94	4.51	3.29	7.80	8.00
$n=4$	6.66	1.42	8.08	6.38	2.33	8.71	8.80
$n=5$	4.43	2.47	6.90	4.01	3.44	7.45	7.80
$n=6$	5.24	2.16	7.40	5.11	2.99	8.10	8.65
$n=7$	3.92	3.08	7.00	3.69	3.61	7.30	7.15
$n=8$	5.38	2.28	7.66	4.67	3.12	7.79	8.20
$n=9$	3.70	3.52	7.22	3.44	3.85	7.29	7.28

charge for Be atoms in $3p$ states shows an inverse even–odd oscillations compared with the HOMO–LUMO energy gap.

Vertical ionization potential, electron affinity and chemical hardness

Chemical hardness has been established as an electronic quantity that, in many cases, may be used to characterize the relative stability of molecules and aggregate through the principle of maximum hardness (PMH) proposed by Pearson [42]. Based on a finite-difference approximation and Koopman's theorem [43], chemical hardness (η) is expressed as:

$$\eta = IP_v - EA_v \quad (7)$$

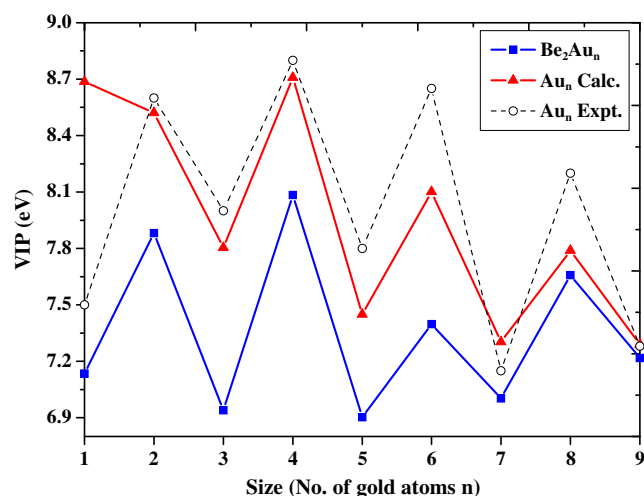


Fig. 4 Calculated and experimentally measured vertical ionization potential for lowest energy structures of Au_{n+2} and Be_2Au_n ($n=1-9$) clusters as a function of cluster size

where IP_v is the vertical ionization potential and EA_v is vertical electron affinity. Thus, the IP_v , EA_v , and η of the most stable Au_{n+2} and Be_2Au_n cluster were calculated and listed in Table 5. One can see that the values of IP_v , EA_v , and η show an obvious oscillating behavior with increasing cluster size. Through the MHP of chemical hardness, such behavior indicates that even-numbered isomers with higher hardness are more stable than the neighboring odd-numbered isomers. This trend agrees well with the above analysis based on HOMO–LUMO energy gaps. Furthermore, the Be_2Au_4 isomer shows the largest chemical hardness and vertical ionization potential of 6.66 eV and 8.08 eV listed in Table 5. The results of the maxima value are also in accord with the averaged atomic binding energy and second-order difference of energy. The relationships of vertical ionization potential and cluster size for Be_2Au_n and Au_{n+2} clusters are plotted in Fig. 4. It is worth noting that the calculated IP_v values of pure gold clusters agree well with experimental data as expected [44]. Unfortunately, there are no experimental IP_v values for Be_2Au_n clusters. Thus, it is hoped that our results will be confirmed by further experimental and theoretical studies.

Conclusions

The geometrical structures, relative stabilities, and electronic properties of small Be_2Au_n ($n=1-9$) clusters were investigated systematically by DFT at the BP86/GENECP level. The results can be summarized as follows.

- (1) The geometrical optimizations indicate that one-Au-atom-capped Be_2Au_{n-1} structure for different sized Be_2Au_n ($n=1-9$) clusters is the dominant growth pattern. The lowest energy isomers of Be_2Au_n clusters have 3D structures at $n=3, 5, 7, 8, 9$. For Be_2Au_2 and Be_2Au_4 clusters, the lowest energy isomers have a planar structure, with structures unlike those of pure gold clusters.
- (2) The averaged atomic binding energies and second-order difference of energies show that the planar Be_2Au_4 isomer is the most stable structure for Be_2Au_n ($n=1-9$) clusters. In addition, the HOMO–LUMO energy gaps show a pronounced even–odd alternation phenomenon as a function of cluster size, showing that clusters with an even number of atoms have enhanced chemical stability.
- (3) NPA results show that the charges in $Be_2Au_{1, 2, 4}$ clusters transfer from Be atoms to the Au frames, whereas in $Be_2Au_{3, 5-9}$, clusters transfer from Au atoms to Be atoms. The internal charge transfers from 2 s states to 2p and 3p states. Vertical ionization

potential and hardness of Be_2Au_n clusters display obvious even–odd alternations, with the maximum occurring at $n=4$.

Acknowledgment The authors are grateful to the National Natural Science Foundation of China (No. 10974138)

References

1. Ferrighi L, Hammer B, Madsen GKH (2009) *J Am Chem Soc* 131:10605–10609
2. Baruah T, Blundell SA, Zope RR (2001) *Phys Rev A* 64:043202–043209
3. Thomas OC, Zheng WJ, Lippa TP, Xu SJ, Lyapustina SA, Bowen KH (2001) *J Chem Phys* 114:9895–9900
4. Fantucci P, Bonačić-Koutecký V, Pewestorf W, Koutecký J (1989) *J Chem Phys* 91:456802
5. Li X, Kuznetsov AE, Zhang HF, Boldyrev AI, Wang LS (2001) *Science* 291:859–861
6. Yoon B, Häkkinen H, Landman U, Wörz AS, Antonietti JM, Abbet S, Judai K, Heiz U (2005) *Science* 307:403–407
7. Pal R, Wang LM, Huang W, Zeng XC (2009) *J Am Chem Soc* 131:3396–3404
8. Eachus RS, Marchetti AP, Muentner AA (1999) *Annu Rev Phys Chem* 50:117–123
9. Li XB, Wang HY, Yang XD, Zhu ZH (2007) *J Chem Phys* 126:084505
10. Scaffardi LB, Pellegrini N, Sanctis Ode, Tocho JO (2005) *Nanotechnology* 16:158–163
11. Häkkinen H, Moseler M, Landman U (2002) *Phys Rev Lett* 89:033401
12. Wang F, Liu P, Zhang DJ (2010) *J Mol Model*. online first. doi:10.1007/s00894-010-0815-4
13. Torres MB, Fernández EM, Balbás LC (2008) *J Phys Chem A* 112:6678–6689
14. Hashmi ASK, Loos A, Littmann A, Braun I, Knight J, Doherty S, Rominger F (2009) *Angew Chem* 351:576–582
15. Chrétien SC, Buratto SK, Metiu H (2007) *Curr Opin Solid State Mater Sci* 11:62–75
16. Autschbach J, Hess BA, Johansson MP, Neugebauer J, Patzschke M, Pykkö P, Reiher M, Sundholm D (2004) *Phys Chem Chem Phys* 6:11–22
17. Shaw CF III (1999) *Chem Rev* 99:2589–2600
18. Albonetti S, Bonelli R, Delaigle R, Femoni C, Gaigneaux EM, Morandi V, Ortolani L, Tiozzo C, Zacchini S, Trifiro F (2010) *Appl Catal A* 372:138–146
19. Neumaier M, Weigend F, Hamper O, Kappes MM (2006) *J Chem Phys* 125:104308
20. Félix C, Sieber C, Harbich W, Buttet J, Rabin I, Schulze W, Ertl G (2001) *Phys Rev Lett* 86:2992–2995
21. Kim SH, Medeiros-Ribeiro G, Ohlberg DAA, Stanley Williams R, Heath JR (1999) *J Phys Chem B* 103:10341–10347
22. Bishea GA, Arrington CA, Behm JM, Morse MD (1991) *J Chem Phys* 95:8765–8778
23. Janssens E, Tanaka H, Neukermans S, Silverans RE, Lievens P (2003) *New J Phys* 5:46.1–46.10
24. Negishi Y, Nakamura Y, Nakamura A, Kaya K (2001) *J Chem Phys* 115:3657–3663
25. Koyasu K, Naono Y, Akutsu M, Mitsui M, Nakajima A (2006) *Chem Phys Lett* 422:62–66

26. Heiz U, Yeretizian C, Stener M, Gisdakis P, Rösch N (1996) *J Chem Phys* 105:5574–5585
27. Heinebrodt M, Malinowski N, Tast F, Branz W, Billas IML, Martin TP (1999) *J Chem Phys* 110:9915–9921
28. Yuan DW, Wang Y, Zeng Z (2005) *J Chem Phys* 122:114310
29. Ghanty TK, Banerjee A, Chakrabarti A (2010) *J Phys Chem C* 114:20–27
30. Guo JJ, Yang JX, Die D (2005) *Phys B* 367:158–164
31. Guo JJ, Yang JX, Die D (2006) *J Mol Struct THEOCHEM* 764:117–121
32. Balducci G, Ciccioni A, Gigli G (2004) *J Chem Phys* 121:7748–7755
33. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millan JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) *Gaussian 03, Revision D.01*. Gaussian Inc, Wallingford, CT
34. Becke AD (1988) *Phys Rev A* 38:3098–3100
35. Perdew JP, Yue W (1986) *Phys Rev B* 33:8800–8802
36. Perdew JP, Wang Y (1992) *Phys Rev B* 45:13244–13249
37. Hay PJ, Wadt WR (1985) *J Chem Phys* 82:299–310
38. Ehlers AW, Bohme M, Dapprich S, Gobbi A, Hollwarth A, Jonas V, Kohler KF, Stegmann R, Veldkamp A, Frenking G (1993) *Chem Phys Lett* 208:111–114
39. Krishnan R, Binkley JS, Seeger R, Pople JA (1980) *J Chem Phys* 72:650–654
40. Barrow RF, Gissane WJG, Travis DN (1965) *Proc R Soc London Ser A* 287:240–251
41. Huber KP, Herzberg G (1979) *Constants of diatomic molecules*. Van Nostrand Reinhold, New York
42. Pearson RG (1997) *Chemical hardness: applications from molecules to solids*. Wiley-VCH, Weinheim
43. Parr RG, Yang W (1989) *Density functional theory of atoms and molecules*. Oxford University Press, New York
44. Jackslath C, Rabin I, Schulze W (1992) *Ber Bunsenges Phys Chem* 96:1200–1204

Receptor recognition mechanism of human influenza A H1N1 (1918), avian influenza A H5N1 (2004), and pandemic H1N1 (2009) neuraminidase

Nipa Jongkon · Chak Sangma

Received: 25 January 2011 / Accepted: 24 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract Influenza A neuraminidase (NA) is a target for anti-influenza drugs. The function of this enzyme is to cleave a glycosidic linkage of a host cell receptor that links sialic acid (Sia) to galactose (Gal), to allow the virus to leave an infected cell and propagate. The receptor is an oligosaccharide on the host cell surface. There are two types of oligosaccharide receptor; the first, which is found mainly on avian epithelial cell surfaces, links Sia with Gal by an α 2,3 glycosidic linkage; in the second, found mainly on human epithelial cell surfaces, linkage is via an α 2,6 linkage. Some researchers believe that NAs from different viruses show selectivity for each type of linkage, but there is limited information available to confirm this hypothesis. To see if the linkage type is more specific to any particular NA, a number of NA-receptor complexes of human influenza A H1N1 (1918), avian influenza A H5N1 (2004), and a pandemic strain of H1N1 (2009) were constructed using homology modeling and molecular dynamics simulation. The results show that the two types of receptor analogues bound to NAs use different mechanisms. Moreover, it was found that a residue unique to avian virus NA is responsible for the recognition of the Sia α 2,3Gal receptor, and a residue unique to human virus NA is responsible for the recognition of Sia α 2,6Gal. We believe that this finding could explain how NAs of different virus origins always possess some unique residues.

Keywords Avian influenza · Hemagglutinin · Neuraminidase · Binding mechanism · Molecular dynamics · Molecular recognition · Glycosidic linkage · Cell receptor

Introduction

To infect a host cell, the influenza virus uses hemagglutinin (HA) to attach to oligosaccharide cell receptors and enter the cell [1]. When the virus is ready to leave to infect another cell, it uses neuraminidase (NA) to cleave the sialic acid–galactose (Sia–Gal) glycosidic linkage in order to be able to exit the cell. As depicted in Fig. 1, pentasaccharides on the cell surface are the main cell receptors for influenza HAs. There are two types of receptors, one with an α 2,3 glycosidic linkage between Sia and Gal, i.e., Sia α 2,3Gal β 1,3GlcNAc β 1,3Gal β 1,4Glc (LSTa), and the other with an α 2,6 linkage between Sia and Gal i.e., NeuAc α 2,6Gal β 1,4GlcNAc β 1,3Gal β 1,4Glc (LSTc). Most of the pentasaccharides found in human epithelial cells have the α 2,6 linkage (referred to as human receptor) while most avian cells have the α 2,3 linkage (referred to as avian receptor). Evidence shows that influenza HAs from different origins have some selectivity for the type of receptors found in their hosts [2–4]. For example, HA from human influenza virus seems to prefer the α 2,6 cell receptor [5, 6]. In contrast, HA from avian influenza seems to prefer the α 2,3 receptor [7–9], and HA from some swine influenza can bind equally to both receptors [10]. These observations are associated with the different types of amino acids near the glycosidic linkage binding site, i.e., residues 222 and 224 (H5 numbering) [11, 12]. This selectivity for host cell receptor according to HA is believed to be a barrier that prevents cross-species infection by the bird flu virus [13]. However, some experiments have

N. Jongkon (✉) · C. Sangma (✉)
Cheminformatics Research Unit, Department of Chemistry,
Faculty of Science, Kasetsart University,
Bangkok, Thailand 10900
e-mail: g5084005@ku.ac.th

C. Sangma
e-mail: fscicsm@ku.ac.th

Fig. 1 **a** Two pentasaccharides on the cell surface representing human cell receptors at the binding conformations for influenza hemagglutinin (HA) from X-ray structures (pdb code: 1JSI [11] and 1RVT [17]). It can be seen that only the three sugar units of both LSTc share a similar conformation. In this study, part of LSTc (Sia α 2, 6Gal β 1, 4GlcNAc) (**b**) was used to represent a human receptor, while part of LSTa (Sia α 2, 3Gal β 1, 3GlcNAc) (**c**) was used for the avian receptor

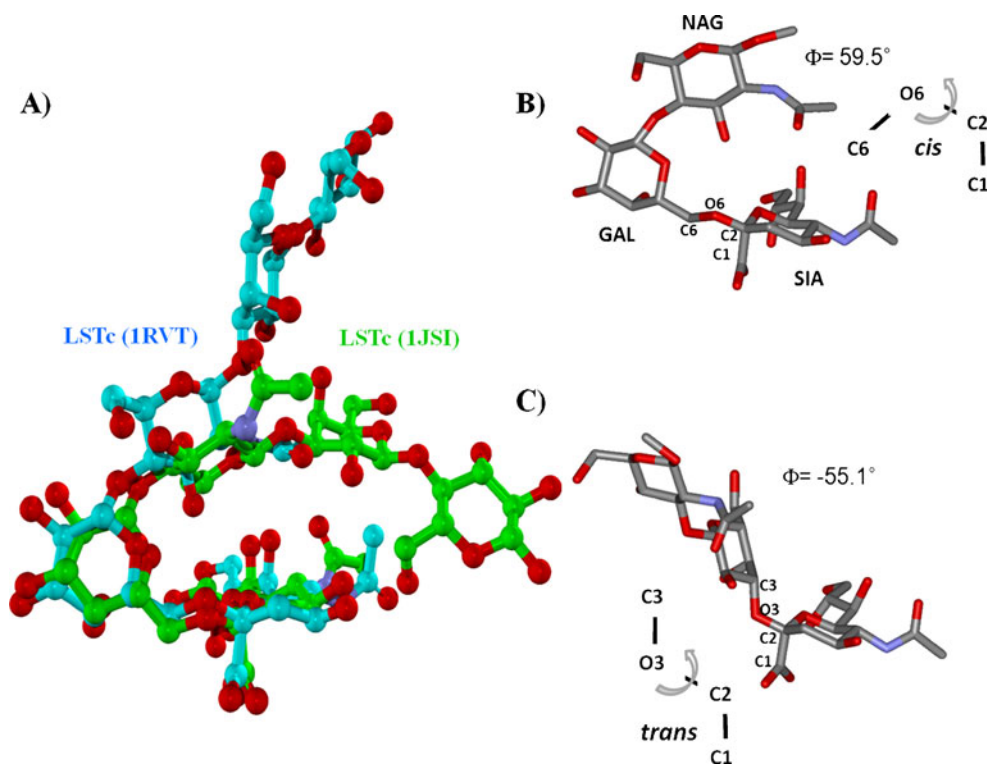


Fig. 2 Superposition of the crystal structures of the Spanish flu 1918 H1N1 NA and avian flu H5N1 NA. The 7 Å residues are shown as *colored lines* (some residues were omitted), and the antiviral drugs Relenza and Oseltamivir are represented by *ball and stick models*. The key residue 347 is shown by a *stick model*. Purple H1N1 NA, green H5N1 NA

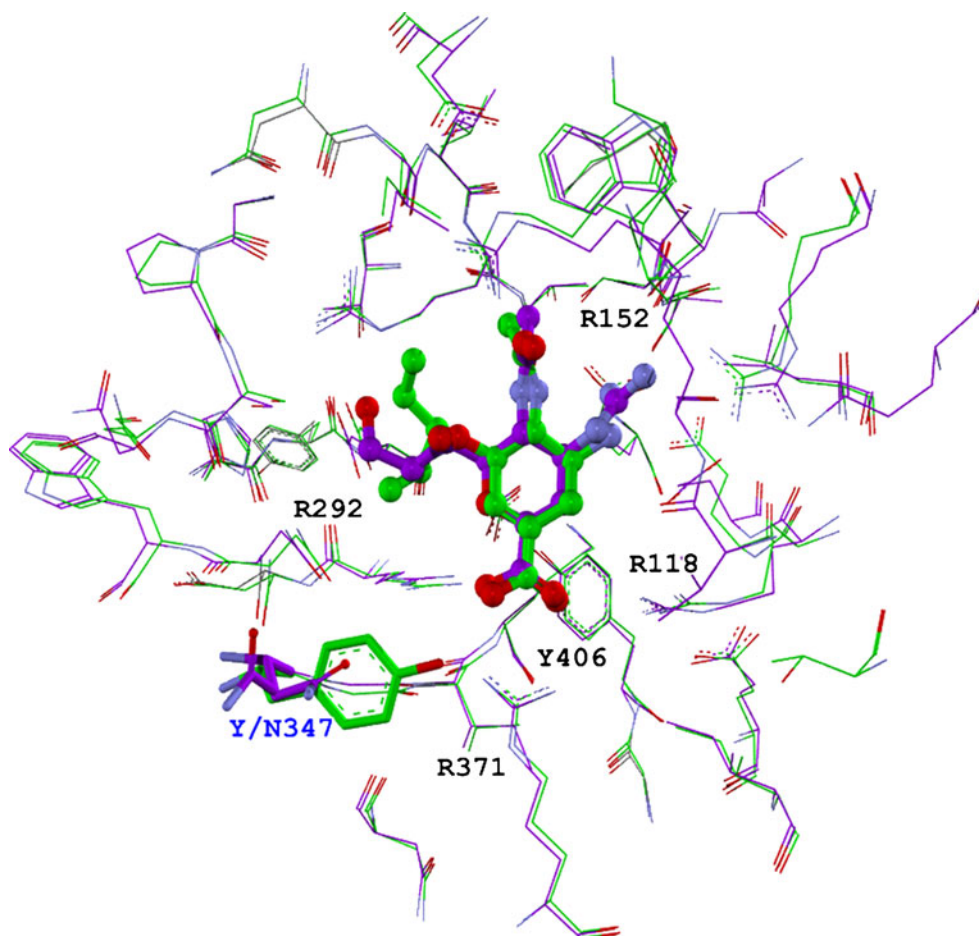
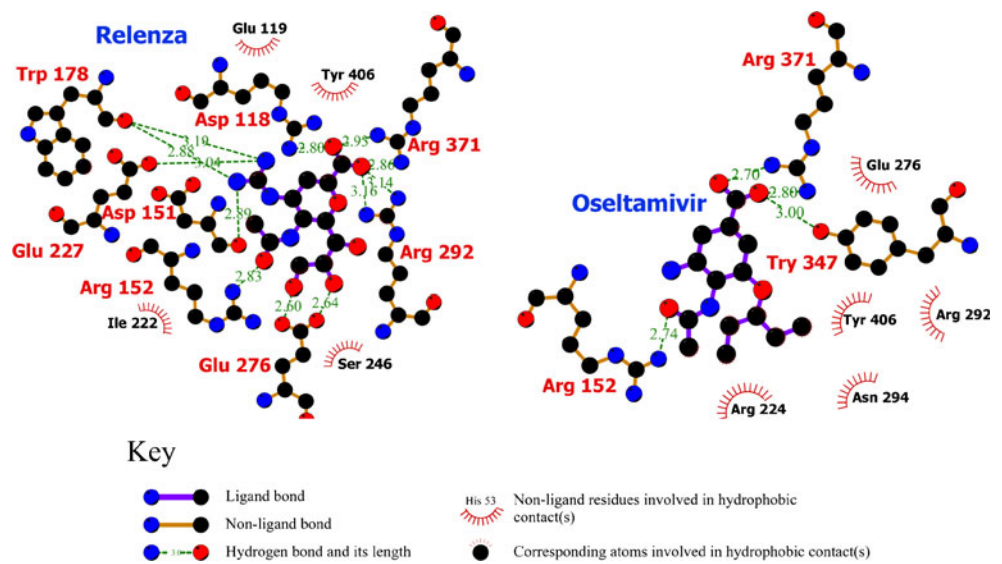


Fig. 3 Atomic contacts of two crystal structures using the Ligplot program [23]: *left* 1918 H1N1 NA complexed with Relenza, *right* 2004 H5N1 NA complexed with Oseltamivir



shown that host selectivity does not rely on HA binding alone, since different influenza strains with common HA have demonstrated different affinity towards the same host [14]. Speculation has arisen that the selectivity process might come from other viral proteins. Among the proteins of influenza A, NA is another protein that also binds to host cell receptors [1]. Many researchers have speculated that the NAs from different influenza virus have a similar selectivity over the two linkage types as found in HA binding [15, 16] but no one has proposed the underlining selectivity mechanisms.

Investigation of their NA sequences and superposition of human and avian influenza viruses has revealed that they have a unique set of residues near their receptor binding sites (RBSs) as shown in Figs. 2–3. In all human H1N1 NA strains, residue 347 is Asp or Asn while in H5N1 NA, this amino acid is Tyr [18]. Some computational approaches have been used to study the important role of this position

[19, 20]. Residue 430 in H1N1 NA is Asn, while it is Arg in H5N1 NA. These residues are not in the Sia binding pocket, and no evidence has been found that they are involved with the binding process. When NA cleaves the Sia–Gal glycosidic linkage, no other part of a cell receptor except Sia can be found within the X-ray structures [21]. Therefore, the RBS of NA is well defined only within the sialic binding area, where several ligands, such as the transition state compound and some enzyme inhibitors, have been crystallized [18, 22].

To determine whether or not these virus-origin dependent residues play any role in NA host cell recognition, models with human and avian receptor analogues, which exhibited the two different types of glycosidic linkage, were constructed. The 2004 influenza virus H5N1, the 1918 influenza virus H1N1, and influenza A H1N1 2009 NA structures were modeled. These models were long enough to cover the points of difference, namely NA residues 347

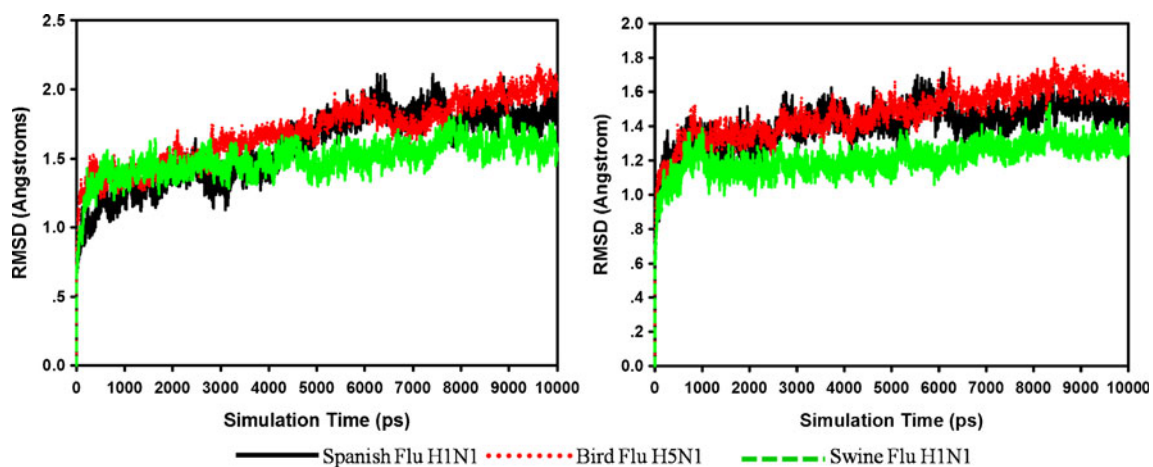


Fig. 4 Backbone atom root-mean-square deviations (RMSD) fluctuation of proteins in complexed with human (*left*) and avian (*right*) receptor

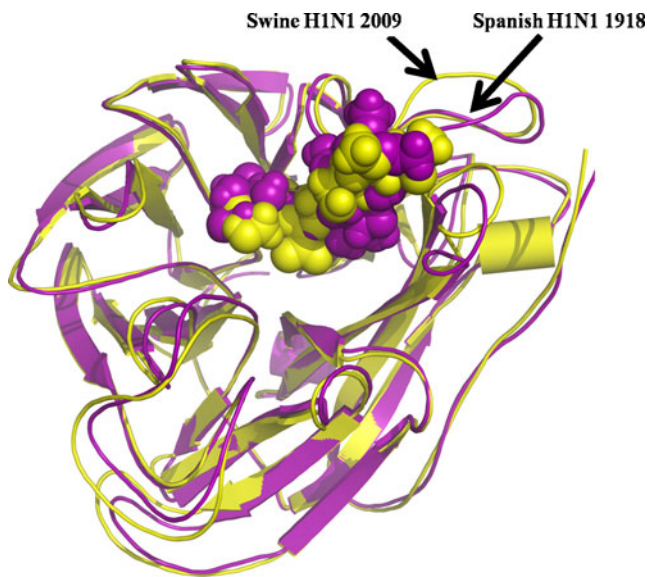


Fig. 5 Superposition of the modeled swine H1N1 2009 (yellow) with the Spanish 1918 H1N1 (purple) derived from the averaged structures of the last 50 ps of molecular dynamics (MD) complexes

and 430, and the binding mechanisms were analyzed and compared. These simulated structures and analysis of their dynamic behavior also provided information on how the hypothetical initial structures, prior to the formation of transition-state complexes, would be affected by the residues next to the Sia binding sites. This information is not available from X-ray structures but is useful in determining how the different glycosidic linkages from human and avian receptors affect NA reaction.

Materials and methods

The binding of human and avian receptor analogues with NA of H5N1 and H1N1 viruses to form complexes where no X-ray structures are available were studied. The receptor models included at least three sugar units to ensure they were long enough to interact with the NA residues of interest.

Preparation of the complex structure

To compare with previous experimental results [24, 25], all simulations used the same protocol previously applied to Spanish 1918 H1N1 influenza [A/Brevig Mission/1/1918 H1N1 strain], avian influenza 2004 H5N1 influenza (A/VietNam/1203/2004(H5N1) and a homology modeled swine 2009 (Sequence ID EPI176472 A/California/04/2009) strain—one of the emerging 2009 strains [26]. In the simulations, the receptor analogues were trisaccharides consisting of the first three sugar units of LSTc and LSTa, i. e., Sia α 2,6Gal β 1,4GlcNAc for the human receptor and Sia α 2,3Gal β 1,3GlcNAc for the avian receptor (refer to Fig. 1b,c). These trisaccharides were capped with an –OME group. Initial NA-receptor complexes were built by docking the receptor analogues into the NA RBS and finding the lowest energy structure from molecular docking using the Lamarckian Genetic Algorithm (LGA) as a searching algorithm with 50 runs for every docking performed by Autodock 4.0 [27]. The complex structures were used as the starting inputs for molecular dynamics (MD) simulations. Two X-ray structures were used as the starting point, i. e., the NA from H5N1 complexed with oseltamivir [22] (PDB code 2HU0) and H1N1 complexed with zanamivir

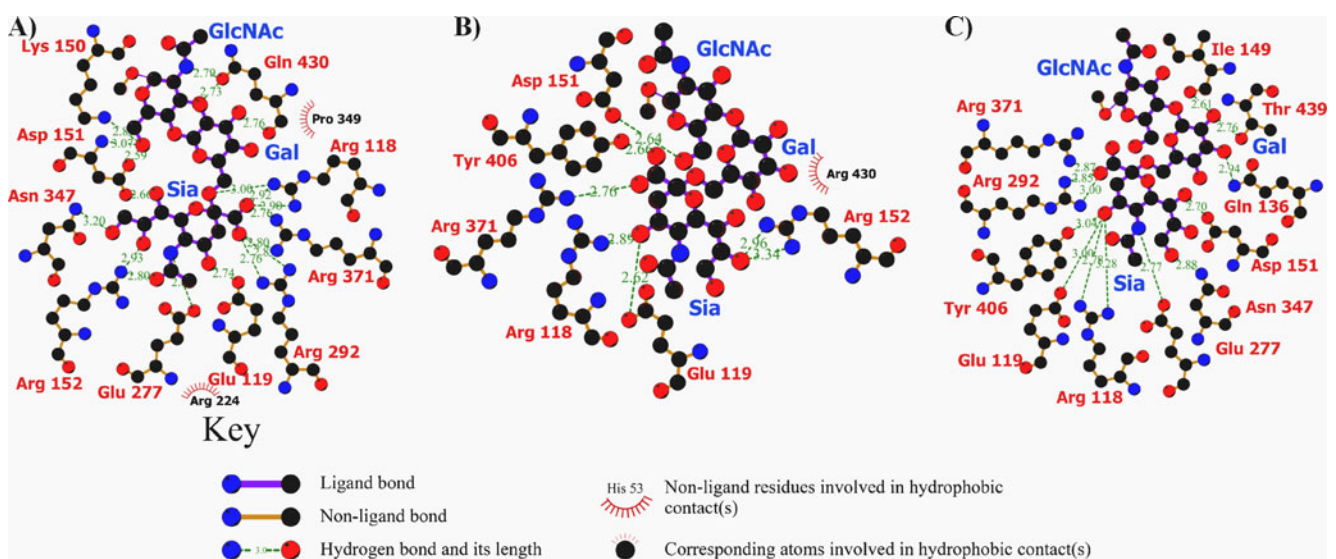


Fig. 6 Atomic contacts between the averaged structures of the 1918 influenza virus H1N1 (a), the 2004 influenza virus H5N1 (b), and influenza A H1N1 2009 NA (c) bound to a human receptor analogue

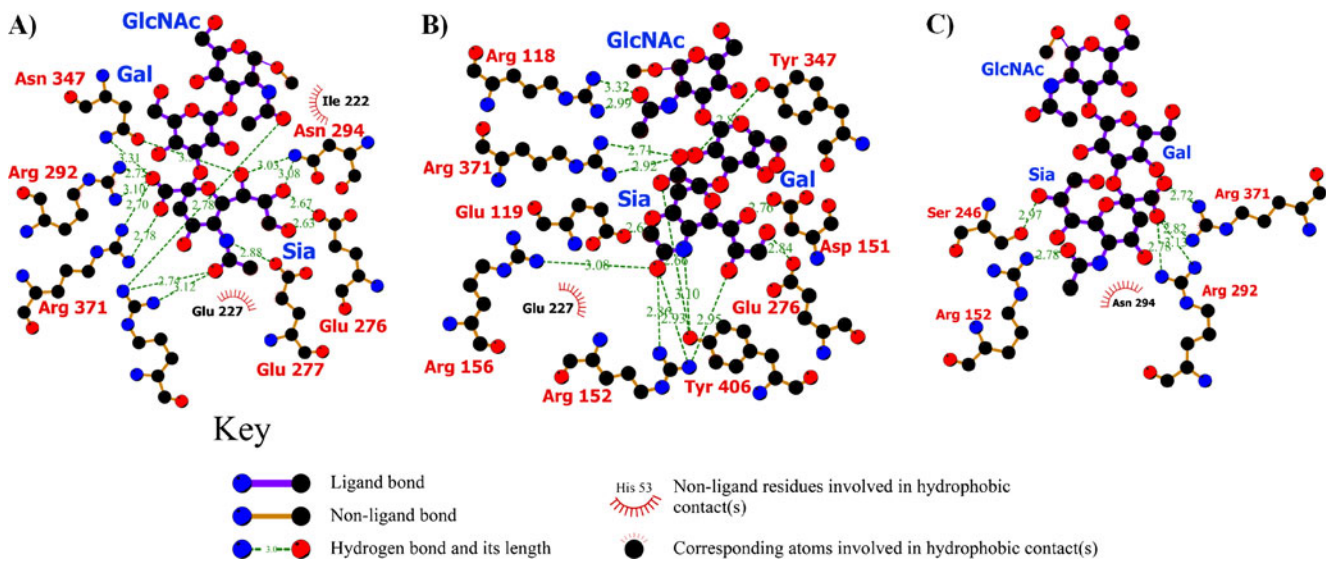


Fig. 7 Atomic contacts between the averaged structures of the 1918 influenza virus H1N1 (a), the 2004 influenza virus H5N1 (b), and influenza A H1N1 2009 NA (c) bound to an avian receptor analogue

Fig. 8 Atomic contacts between the averaged structure of the avian flu 2004 H5N1 NA bound to an avian receptor analogue (colored surfaces are based on calculated charges from Accelrys Discovery studio 2.5)

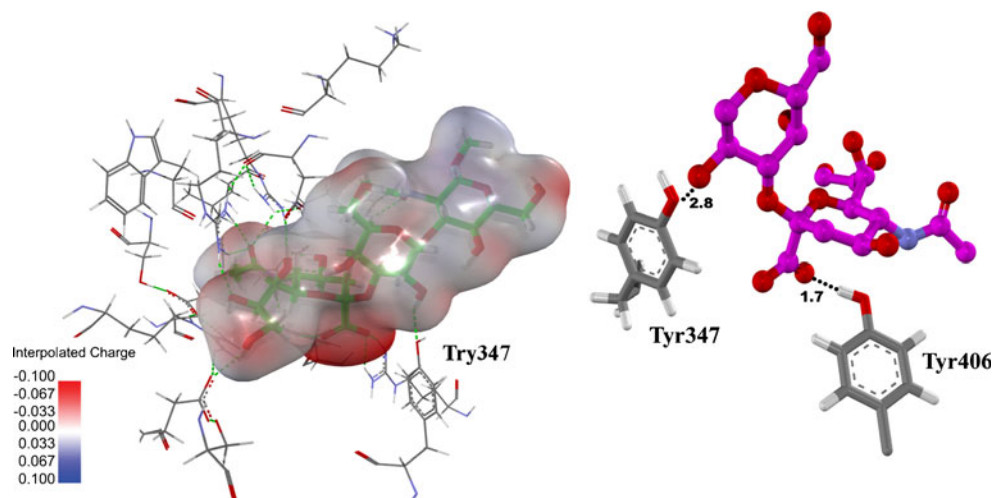


Table 1 Summary of all residues involved in the recognition process

Neuraminidase type	Receptor type	Sialic acid recognition	Galactose recognition	GlcNAc recognition
Spanish flu 1918	Human	R118 ^a ,E119,D151,R152,E277, R292 ^a ,N347,R371 ^a	Q430	K150,D151, Q430
	Avian	R152,E276,E277 ^a ,R292 ^a ,N294 ^a , N347,R371 ^a		
Avian flu 2004	Human	R118 ^a ,E119,R152 ^a ,R371 ^a ,Y406 ^a		D151
	Avian	E119,D151,R152,R156,R371 ^a , Y406 ^a	Y347	R118
Swine flu 2009	Human	R118 ^a ,E119,D151,E277 ^a ,R292 ^a , N347,R371 ^a ,Y406	Q136,I149,T439	
	Avian	R152,S246,R292 ^a ,R371 ^a		

^a >50 % hydrogen bond occupancy

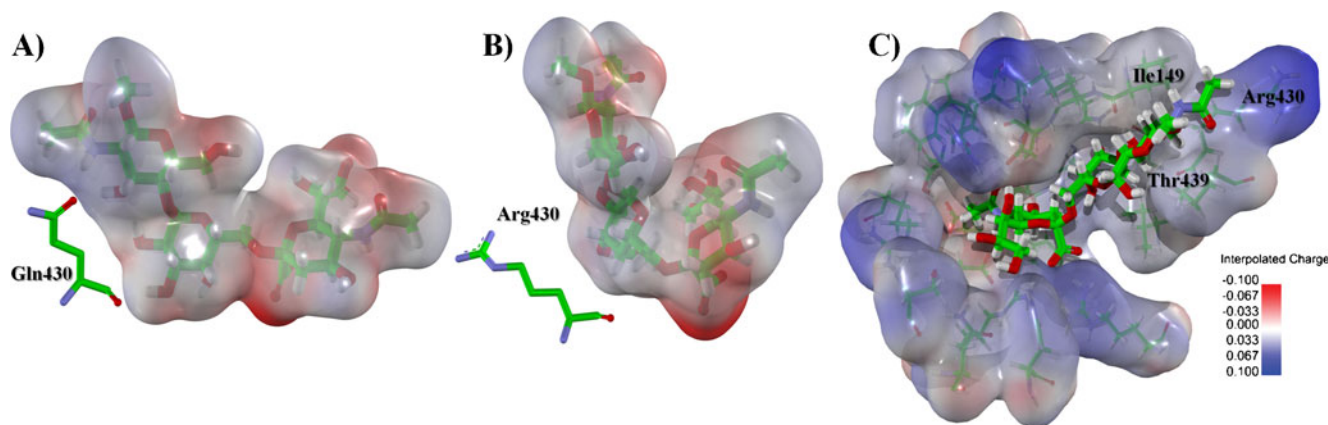


Fig. 9 Atomic contacts between the averaged structures of **a** Spanish flu, **b** avian flu 2004 H5N1 NA and **c** swine flu bound to a human receptor analogue (*colored surfaces* are based on calculated charges from Accelrys Discovery studio 2.5)

(PDB code 3B7E) [18]. MD simulations were undertaken until the complexes reached an equilibrium state.

Molecular dynamics conditions

For the MD runs, Glycam06 parameters [28] and the AMBER force field [29] were used for the sugar receptors and proteins, respectively. Hydrogen atoms and counter Na^+ or Cl^- ions were added to neutralize protein structures taken from X-ray structures using the LEAP module. Initial complex structures were first solvated using the TIP5P water model [30, 31]. About 7,000 water molecules in a truncated octahedron box were systematically relaxed, while keeping the proteins and receptors restrained with forces of $500 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. The whole systems were relaxed at 0 K and then heated from 0 to 300 K with 10 kcal mol^{-1} harmonic restraints for 100 ps, with NVT equilibration where bonds involving hydrogen are constrained using the SHAKE

algorithm [32] and with a 9-Å non-bonded cutoff. The production runs of 10 ns at 300 K using the SANDER module were then performed. The structures started to be stable after 1 ns with some fluctuations, and were used to analyze both structural changes and fluctuations as illustrated in Figs. 4 and 5. Xmgrace <http://plasma-gate.weizmann.ac.il/Grace/>, LIGPLOT program [23], Accelrys Discovery Studio 2.5 (Accelrys, San Diego, CA), PyMol 0.99 program [33] and AMBER tools were used to visualize and manipulate images.

Results and discussion

Within the Sia binding region, all the simulated structures showed binding patterns resembling the transition-state [21] and the analogue drugs [18, 22] as shown in Figs. 6 and 7. This implied that Sia binding was still the most important

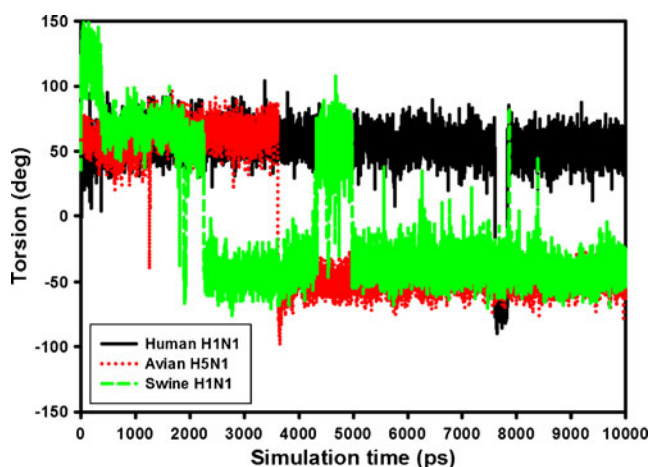


Fig. 10 Variation in the torsion angle for the human receptor complexed with different NAs

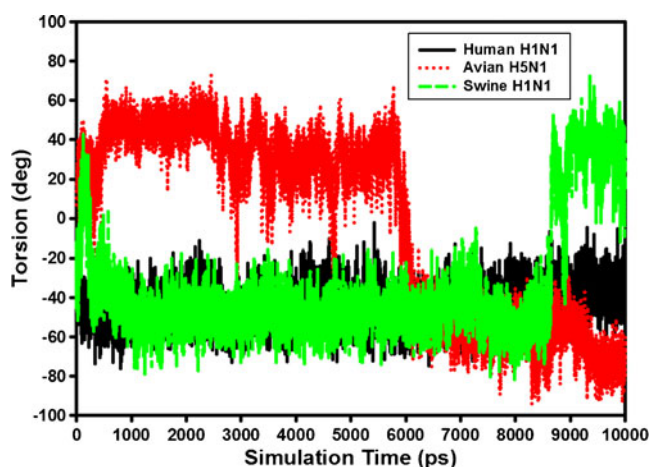


Fig. 11 Variation in the torsion angle for the avian receptor complexed with different NAs

part of all NAs regardless of whether the substrates were from human or avian cell receptors. However, three additional features of Sia binding were observed that have not been previously reported.

Arginine triad

First, in all H5N1 and H1N1 NA binding, the avian receptor weakened some interactions in the binding pocket. Second, the Sia binding of the human receptor seemed to rely on several NA residues, while binding of the avian receptor depended on fewer residues. Third, besides the same interactions found in the drug-NA complexes, an additional interaction between Tyr406 of H5N1 NA and Sia was observed as illustrated in Fig. 8. These features had a slight effect on Sia binding in both NA-cell receptor systems, but had a greater effect on Gal binding.

Apart from the Sia binding area, the other sugars started to show some specific interaction to each linkage. There were also other newly observed interactions among the NA systems. In the system of the avian receptor analogue bound to H5N1 NA, the Gal hydroxyl groups O3 and O4 formed hydrogen bonds with Tyr347, but none were observed with the avian receptor bound to H1N1 NA as shown in Fig. 7 and Table 1. However, a similar Gal interaction to the amino acid side chain of H1N1 NA was seen when the receptor analogue was human type, between residue Gln430 of Spanish H1N1 NA and O3 or O4 of α 2,6-Gal (shown in Fig. 9a). The number of residues involved in Gal binding was increased in swine H1N1 NA and included Gln136, Ile149 and Thr439, as seen in Figs. 6 and 9c.

Receptor binding conformations

For modeling, the two receptors had their binding characteristics expressed in terms of their binding conformations.

According to the torsion angle as defined in Fig. 1b,c monitored throughout each simulation, which reflected the *cis* or *trans* conformation of the receptor, three behaviors were found. In the first case, the torsion angle kept its initial value until the end of the simulation. In the second case, there was a change in its value after some time, but it remained constant thereafter. In the third case, the torsion angle fluctuated between two values (Figs. 10, 11).

Human and avian receptor recognition

It can be concluded from these results that different modes of receptor–protein recognition occur depending on the type of receptor and origin of the virus. In H5N1 NA, the enzyme uses the Tyr406 hydroxyl hydrogen to form a bond with the Sia carboxylate group, while in the human virus NA, the Tyr406–Sia hydrogen bond was not observed. In other words, the interaction was found only in the H5N1 NA-receptor complex, so we believe that this should be a part of the avian receptor recognition process of the avian influenza NA. However, this recognition process could not be observed after the enzymatic reaction had taken place and the conformation of Sia had changed. Another specific recognition process was observed at Gal. In all NAs, either the O3 or O4 of the Gal hydroxyl of the receptor was observed to form hydrogen bonds with the NA side chains. However, the recognition was observed in H5N1 NA only when the receptor was avian type, while with both H1N1 NAs (human and swine), this interaction was observed only when the receptor was human type.

In other words, the avian flu virus NA would recognize Gal only when the receptor was avian type, and the human or swine flu virus would recognize Gal only when the receptor was human type. Tyr347, found in all H5N1 NAs, was seen to be responsible for this recognition in avian influenza NA [34]. In Spanish H1N1 NA, the recognition

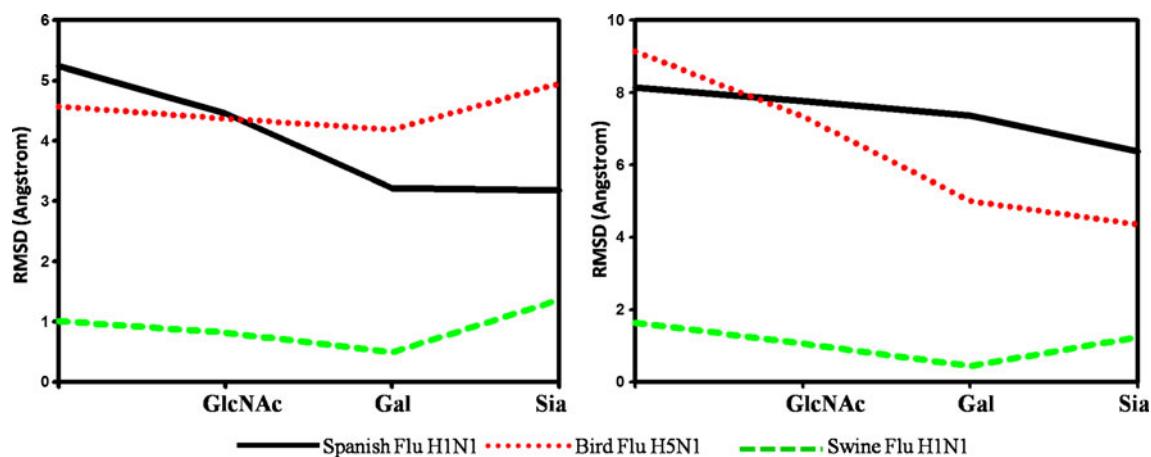


Fig. 12 RMSD of each sugar residue of human receptor (*left*) and avian receptor (*right*)

was due to Gln430, and in swine H1N1 NA, to Gln136, Ile149 and Thr439. The interactions observed in the simulations are summarized in Table 1.

Although more interactions were found within the Sia binding pocket than within the Gal binding regions, the residue fluctuation of Gal was lowest as seen in Fig. 12. It is possible that, in all systems, bound Gal was in the optimum conformation for this sugar in terms of receptor analogue structure. That is, the Gal binding regions were built to accommodate the optimum binding conformation of the receptor analogues and very little change was needed. Besides an easily observed recognition process arising from electrostatic interaction, this means that Gal was recognized by its shape although some Gal movement was still allowed. When recognition occurs, it appears that substrate fixing is enhanced, making the conformation of the cell receptor more rigid. If this is true, the rate of NA reaction will be determined by the molecular recognition process due to some specific residues.

Conclusions

The main contribution to all NA bindings in this study was within the Sia binding region, but differences between each receptor type binding pattern were observed outside this area. Receptor–NA recognition mechanisms were proposed from the simulations performed, and used to study the role of the unique residues found in particular NAs of different origin. As a result, the different NA residues bound to Gal made the glycosidic linkage conformation for the enzyme substrates appear different. The distinguished conformations allowed the Gals of avian and human receptor—the leaving groups of the NA reaction—start to leave and yield two different products. Hence, different rates of reaction would be expected. Additionally, it was found that, within the Sia binding region, NA-receptor recognition depended on similar sets of amino residues (Table 1) in terms of hydrogen bonding with the receptor. However, within the Gal binding region, avian virus H5N1 NA used Tyr347 to recognize Gal of the avian receptor while human virus NA used Gln430. This means that these unique residues, Tyr 347 of H5N1 and Gln430 of human virus NA, are important in the initial complex formation process of the NA-substrate. In other places where hydrogen bonds were not frequently observed, the recognition process occurred by shape as in the “lock-and-key” model. In the latter cases, most of the binding pocket residues had to be considered as equally important. It should be noted that no conclusion could be reached as to how the recognition process affects the complete receptor–NA reaction. The proposed recognition process and binding mechanisms were those of the initial avian and human receptor–NA complexes. There are

other reaction steps and intermediates following this step that have to be considered in detail.

Acknowledgments This work was supported by the Commission on Higher Education (CHE-PhD-SW_20060163), Thailand. A Scholarship from The Graduate School of Kasetsart University and Bilateral Research Cooperation (BRC2/2551) from the Faculty of Science, the National Research University Project and the postgraduate education and the Thailand Research Fund (DBG528004) are gratefully acknowledged for partially financial support. The authors would like to thank Mr. Wilhelm Josef. Holzschuh, A. Univ. Prof. Dr. Peter Wolschann, Assoc. Prof. Dr. Supa Hannongbua, and Dr. Matthew Paul Gleeson for helpful comments. The authors acknowledged the Thai National Grid Project (ThaiGrid) and WATA cluster faculty of Engineering, Kasetsart University (URL: <http://www.eng.ku.ac.th>) for providing computing resources that contributed to the research results reported within this paper.

References

1. Wagner R, Matrosovich M, Klenk HD (2002) Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Rev Med Virol* 12:159–166. doi:10.1002/rmv.352
2. Suzuki Y (2005) Sialobiology of influenza: molecular mechanism of host range variation of influenza viruses. *Biol Pharm Bull* 28:399–408
3. Gambaryan A, Webster R, Matrosovich M (2002) Differences between influenza virus receptors on target cells of duck and chicken. *Arch Virol* 147:1197–1208. doi:10.1007/s00705-002-0796-4
4. Gambaryan AS, Karasin AI, Tuzikov AB, Chinarev AA, Pazynina GV, Bovin NV, Matrosovich MN, Olsen CW, Klimov AI (2005) Receptor-binding properties of swine influenza viruses isolated and propagated in MDCK cells. *Virus Res* 114:15–22
5. Connor RJ, Kawaoka Y, Webster RG, Paulson JC (1994) Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology* 205:17–23
6. Glaser L, Conenello G, Paulson J, Palese P (2007) Effective replication of human influenza viruses in mice lacking a major α 2,6 sialyltransferase. *Virus Res* 126:9–18
7. Gambaryan AS, Tuzikov AB, Pazynina GV, Webster RG, Matrosovich MN, Bovin NV (2004) H5N1 chicken influenza viruses display a high binding affinity for Neu5Ac α 2-3Gal β 1-4 (6-HSO3)GlcNAc-containing receptors. *Virology* 326:310–316
8. Gambaryan A, Yamnikova S, Lvov D, Tuzikov A, Chinarev A, Pazynina G, Webster R, Matrosovich M, Bovin N (2005) Receptor specificity of influenza viruses from birds and mammals: new data on involvement of the inner fragments of the carbohydrate chain. *Virology* 334:276–283
9. Li M, Wang B (2006) Computational studies of H5N1 hemagglutinin binding with SA- α -2, 3-Gal and SA- α -2, 6-Gal. *Biochem Biophys Res Commun* 347:662–668
10. Ito T, Couceiro JNSS, Kelm S, Baum LG, Krauss S, Castrucci MR, Donatelli I, Kida H, Paulson JC, Webster RG, Kawaoka Y (1998) Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. *J Virol* 72:7367–7373
11. Ha Y, Stevens DJ, Skehel JJ, Wiley DC (2001) X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. *Proc Natl Acad Sci USA* 98:11181–11186. doi:10.1073/pnas.201401198
12. Vines A, Wells K, Matrosovich M, Castrucci MR, Ito T, Kawaoka Y (1998) The role of influenza A virus hemagglutinin residues 226 and 228 in receptor specificity and host range restriction. *J Virol* 72:7626–7631

13. Rogers GN, Paulson JC (1983) Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology* 127:361–373
14. Mitnaul LJ, Matrosovich MN, Castrucci MR, Tuzikov AB, Bovin NV, Kobasa D, Kawaoka Y (2000) Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. *J Virol* 74:6015–6020. doi:10.1128/jvi.74.13.6015-6020.2000
15. Katinger D, Mochalova L, Chinarev A, Bovin N, Romanova J (2004) Specificity of neuraminidase activity from influenza viruses isolated in different hosts tested with novel substrates. *Arch Virol* 149:2131–2140. doi:10.1007/s00705-004-0364-1
16. Shtyrya Y, Mochalova L, Voznova G, Rudneva I, Shilov A, Kaverin N, Bovin N (2009) Adjustment of receptor-binding and neuraminidase substrate specificities in avian–human reassortant influenza viruses. *Glycoconj J* 26:99–109. doi:10.1007/s10719-008-9169-x
17. Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, Vasisht N, Steinhauer DA, Daniels RS, Elliot A, Wiley DC, Skehel JJ (2004) The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* 303:1838–1842. doi:10.1126/science.1093155
18. Xu X, Zhu X, Dwek RA, Stevens J, Wilson IA (2008) Structural characterization of the 1918 influenza virus H1N1 neuraminidase. *J Virol* 82:10493–10501. doi:10.1128/jvi.00959-08
19. Rungrotmongkol T, Udommaneethanakit T, Malaisree M, Nunthaboot N, Intharathep P, Sompompisut P, Hannongbua S (2009) How does each substituent functional group of oseltamivir lose its activity against virulent H5N1 influenza mutants? *Biophys Chem* 145:29–36
20. Rungrotmongkol T, Intharathep P, Malaisree M, Nunthaboot N, Kaiyawet N, Sompompisut P, Payungporn S, Poovorawan Y, Hannongbua S (2009) Susceptibility of antiviral drugs against 2009 influenza A (H1N1) virus. *Biochem Biophys Res Commun* 385:390–394
21. <http://www.pdb.org/pdb/explore/explore.do?structureId=2BAT>
22. Russell RJ, Haire LF, Stevens DJ, Collins PJ, Lin YP, Blackburn GM, Hay AJ, Gamblin SJ, Skehel JJ (2006) The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* 443:45–49. http://www.nature.com/nature/journal/v443/n7107/suppinfo/nature05114_S1.html
23. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* 8:127–134. doi:10.1093/protein/8.2.127
24. Auewarakul P, Suptawiwat O, Kongchanagul A, Sangma C, Suzuki Y, Ungchusak K, Louisirirochanakul S, Lerdsamran H, Pooruk P, Thitithanyanont A, Pittayawonganon C, Guo CT, Hiramoto H, Jampangern W, Chunsutthiwat S, Puthavathana P (2007) An avian influenza H5N1 virus that binds to a human-type receptor. *J Virol* 81:9950–9955
25. Jongkon N, Mokmak W, Chuakheaw D, Shaw PJ, Tongsimma S, Sangma C (2009) Prediction of avian influenza A binding preference to human receptor using conformational analysis of receptor bound to hemagglutinin. *BMC Genomics* 10 (Suppl 3)
26. Kumar S, Chusid M, Willoughby R, Havens P, Kehl S, Ledebor N, Li S-H, Henrickson K (2009) Introduction of novel swine-origin influenza A (H1N1) virus into Milwaukee, Wisconsin in 2009. *Viruses* 1:72–83
27. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 19:1639–1662. doi:10.1002/(sici)1096-987x(19981115)19:14<1639::aid-jcc10>3.0.co;2-b
28. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ (2008) GLYCAM06: a generalizable biomolecular force field. *Carbohydrates J Comput Chem* 29:622–655. doi:10.1002/jcc.20820
29. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688. doi:10.1002/jcc.20290
30. Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935. doi:citeulike-article-id:297084
31. Mahoney MW, Jorgensen WL (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J Chem Phys* 112:8910–8922
32. Ryckaert J, Ciccotti G, Berendsen H (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23:327–341. doi:citeulike-article-id:2565765
33. Delano WL (2002) The PyMOL Molecular Graphics System. <http://www.pymol.org>
34. Yu K, Luo C, Qin G, Xu Z, Li N, Liu H, Shen X, Ma J, Wang Q, Yang C, Zhu W, Jiang H (2009) Why are oseltamivir and zanamivir effective against the newly emerged influenza A virus (A/H1N1)? *Cell Res* 19:1221–1224

Structural determinants of benzodiazepinedione/peptide-based p53-HDM₂ inhibitors using 3D-QSAR, docking and molecular dynamics

Fangfang Wang · Yan Li · Zhi Ma · Xia Wang ·
Yonghua Wang

Received: 17 December 2010 / Accepted: 14 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract As a tumor suppressor, p53 protein regulates the cell cycle and is involved in preventing tumorigenesis. The protein level of p53 is under the tight control of its negative regulator human double minute 2 (HDM₂) via ubiquitination. Therefore, the design of inhibitors of HDM₂ has attracted much interest of research on developing novel anticancer drugs. Presently, two classes of molecules, i.e., the 1,4-benzodiazepine-2,5-diones (BDPs) and N-Acylpolyamine (NAPA) derivatives were studied by three-dimensional quantitative structure–activity relationship (3D-QSAR) modeling approaches including the comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) as promising p53-HDM₂ inhibitors. Based on both the ligand-based and receptor-guided (docking) alignments, two optimal 3D-QSAR models were obtained with good predictive power of $q^2=0.41$, $r^2_{\text{pred}}=0.60$ for BDPs, and $q^2=0.414$, $r^2_{\text{pred}}=0.69$ for NAPA analogs, respectively. By analysis of the model and its related contour maps, it is revealed that the electrostatic interactions contributed much larger to the compound binding affinity

than the steric effects. And the contour maps intuitively suggested where to modify the molecular structures in order to improve the binding affinity. In addition, molecular dynamics simulation (MD) study was also carried out on the dataset with purpose of exploring the detailed binding modes of ligand in the HDM₂ binding pocket. Based on the CoMFA contour maps and MD-based docking analyses, some key structural aspects responsible for inhibitory activity of these two classes of compounds were concluded as follows: For BDPs, the R₁ and R₃ regions should have small electronegativity groups; substituents R₂ and R₄ should be larger, and R₃ substituent mainly involves in H-bonds forming. For NAPA derivatives, bulky and electropositive groups in ring B and ring A, small substituent at region P is favorable for the inhibitory activity. The models and related information, we hope, may provide important insight into the inhibitor-p53-HDM₂ interactions and be helpful for facilitating the design of novel potent inhibitors.

Keywords p53 · HDM₂ · 3D-QSAR · CoMFA · CoMSIA · Molecular docking · Molecular dynamics

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1041-4) contains supplementary material, which is available to authorized users.

F. Wang · Z. Ma · X. Wang · Y. Wang
Bioinformatics Center, Northwest A&F University,
Yangling, Shaanxi 712100, China

Y. Li
School of Chemical Engineering,
Dalian University of Technology,
Dalian, Liaoning 116024, China

Y. Wang (✉)
College of Life Sciences, Northwest A&F University,
Yangling, Shaanxi 712100, China
e-mail: yh_wang@nwsuaf.edu.cn

Introduction

The tumor suppressor p53 protein is a transcription factor which involves in controlling the cell cycle and monitoring the integrity of the genome which has made it known as the “guardian of the genome” [1]. Under non-stressed conditions, p53 has a short half-life [2] and maintains in a low level. While under stresses, such as DNA damage and abnormal growth regulation, p53 will be activated and initiates a succession of events which result in growth arrest or apoptosis leading to the elimination of genetically altered cells [3]. Because of its significant roles in keeping the

integrity of genome, it has become a main target of much tumor therapeutics. The loss of p53 functions in many cancers is a result of disabling p53 through mutation or alterations of various components of the pathways that regulate p53. Evidence has shown that p53 mutates in ~50% of human cancers, and in the remaining of human cancers, p53 is in the wild-type but dysfunctional situation [4, 5], due to the precise control of its negative regulator HDM₂ in cells [6, 7].

HDM₂ is a ring finger protein, which can form autoregulatory feedback loop with p53 protein [8, 9]. On one hand, HDM₂ binds to the N-terminal transactivation domain of p53 to block p53 transcriptional activity [10, 11]. On the other hand, HDM₂ exports the complex from the nucleus to the cytoplasm to promote its degradation. More importantly, HDM₂ serves as an E₃ ligase to target p53 for degradation through the proteasome pathway [12]. The detailed interaction mode between HDM₂ and p53 has been revealed by the X-ray crystallography, which structure shows that HDM₂ has a large hydrophobic cleft to bind to α -helix of p53 protein, involving three hydrophobic residues (Phe19, Trp23, and Leu26) in p53 [11], which enhances the stability of the binary complex.

Recently, several approaches have been developed to activate or restore the activity of p53, for example, the activation of wild-type p53 can be made through drugging its modification proteins, stimulating family member p63/p73, or disturbing interactions between HDM₂ and wild-type p53 [13]. And the latter case has allowed for successful identification of many HDM₂-p53 complex inhibitors for cancer treatment [14–16], such as the analogs of cis-imidazole [14], spiro-oxindole [17], benzodiazepinedione [18], quilinol [19] and terphenyl derivatives [20].

By now, the most potent and well-characterized p53-HDM₂ inhibitors are nutlins identified by high-throughput screening method, and the second most potent series are benzodiazepinediones (BDPs) [21]. These inhibitors showed high binding affinity to p53-HDM₂ complex and desirable pharmacokinetic profiles in cells [22]. N-Acylpolyamine analogs (NAPA), another class of peptide inhibitors, have been proved as potent antagonists of the HDM₂-p53 interaction in vitro and in cell-based assays [18, 23]. NAPA shows characteristics of cyclic peptides in β -sheet conformations, and these helical β -peptides can be chemically modified to bind to HDM₂ with good affinity.

Clearly, in vitro assessment of the activity of p53-HDM₂ inhibitors remains a labor-intensive and time-consuming operation. Therefore, more efficient and economical alternative methods should be employed, such as in silico molecular modeling approaches that are used for the purpose of predicting and prioritizing chemicals for subsequent in vitro and in vivo screening. 3D-QSAR (quantitative structure-activity relationship) method, as effective computational

method, has been widely used as a way to find out various interactive fields making impact on the activity and thus to help forecasting and designing of novel inhibitors [24–27, 33, 35], as well as the HDM₂-p53 inhibitors, including Isoindolinone [28], Terphenyls [20] and Chalcones Derivatives [29]. To our knowledge, there is still a lack of in silico modeling of these BDPs and NAPA analogs of the complex inhibitors.

BDPs and NAPA analogs have been reported as promising HDM₂-p53 interaction inhibitors [18, 30–32]. BDPs bound to HDM₂ in the p53-binding cavity in micromolar concentration, and the most potent compound 41 possessed very strong anti-HDM₂ activity with an IC₅₀ of 0.25 μ M. Recent evidence showed that the structures of this class of inhibitors had four available sites, i.e., R₁, R₂, R₃ and R₄ shown in Table S1 for modifications, thus improving inhibitory activity and cell permeability for a candidate [18, 30, 31]. For peptide-based NAPA analogs, a synthesized single trimeric inhibitor showed very potent inhibition on HDM₂ protein, and the most potent compound 102 inhibits HDM₂ binding to p53 with IC₅₀ value of 2.3 μ M [32]. Traditionally, large molecules would be difficult to produce in sufficient quantities for drug development; however, NAPA derivatives offer the possibility of using a short oligomeric scaffold to develop potent p53-HDM₂ inhibitors. Although these two classes of inhibitors are important for treatment of cancer, they have not received much attention from a theoretical perspective so far. Therefore, in this paper, these two classes of inhibitors were analyzed using theoretical computations, with purpose to build related 3D-QSAR models by comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) approaches [33]. In addition, molecular docking and molecular dynamics were also performed to further understand the structure properties and the probable binding modes of these inhibitors at the allosteric site of the receptors. Evidence has proved that this was the first time on 3D-QSAR modeling of these two classes of p53-HDM₂ inhibitors. All these methods applied on the inhibitors could not only help understanding ligand–receptor interactions but also provide useful and rational suggestions for further design of new drug candidates of p53-HDM₂ for cancer therapy.

Methods and materials

Data sets and biological activity

A total of 108 molecules with available IC₅₀ were collected from the references [18, 30–32], these molecules comprised two different classes of diverse 1,4-benzodiazepine-2,5-diones (BDPs) and N-Acylpolyamine (NAPA) derivatives, whose IC₅₀ values were converted into pIC₅₀ ($-\log$ IC₅₀)

values (shown in Tables S1–S2) and which were used as dependent variables in 3D-QSAR analysis. Division of the training and test sets (Tables S1–S2 marked with ^a) was carried out by considering that the test compounds should represent the structural diversity and a range of biological activities similar to that of the training set. All molecular modeling calculations were performed using SYBYL package (Tripos Associates, St. Louis, MO), where Gasteiger–Marsili charges for BDPs and Gasteiger–Hückel charges for NAPA respectively were added and energy minimization for each molecule was performed by Tripos force field [34] with the Powell conjugate gradient minimization algorithm. The minimization wouldn't be finished until the energy gradient convergence criterion of 0.05 kcal mol⁻¹ Å⁻¹ was reached.

Conformational sampling and alignment

For both the CoMFA and CoMSIA studies, the most important step of modeling is that all molecular structures are aligned to a suitable conformational template, which is assumed to be the most potent conformation. In this work, we applied two alignment methods, the first one of which denoted by ligand-based alignment selected the most potent inhibitor for each class (compounds 41, 102, respectively) as the template. All inhibitors for each class in the data set were then aligned to a common substructure (shown in Fig. 1) of the template using the “align database” command in SYBYL software to produce valid and reliable CoMFA and CoMSIA models. The second alignment method we employed was the receptor-based alignment, where all molecules were docked into the receptors and then the top scored conformations were further used directly for CoMFA and CoMSIA analysis. During this process, we not only explored the traditional docking alignment but also a novel receptor-guided consensus dynamics alignment for comparison of the docking

difference between them. Finally, the conformations derived from the dynamics alignment, which is the optimal one we found, was applied for construction of the 3D-QSAR models. Figures 2 and 3 show the aligned conformations derived from the two methods.

3D-QSAR analysis

CoMFA and CoMSIA studies were carried out with the 3D cubic lattice with grid spacing of 2 Å in x, y and z directions, using an sp³ carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0. Cutoff values for both steric and electrostatic fields were set to 30 kcal mol⁻¹. The CoMSIA models were based on the molecular similarity indices with the same lattice box used for the CoMFA calculations. However, it had several advantages over CoMFA such as greater robustness regarding both region shifts and small shifts within the alignments [35]. Other than steric, electrostatic fields described in the CoMFA method, five physicochemical properties: steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D), and hydrogen bond acceptor (A) were evaluated using the standard settings: probe with charge +1, radius 1 Å and hydrophobicity +1, hydrogen-bond donating +1, hydrogen-bond accepting +1, attenuation factor of 0.3 and grid spacing 2 Å. CoMSIA similarity indices ($A_{F,k}^q(j)$) for a molecule j with atoms i at a grid point q are calculated by Eq. 1:

$$A_{F,k}^q(j) = - \sum \omega_{\text{probe},k} \omega_{ik} e^{-\alpha r} \quad (1)$$

In this equation, k represents five physicochemical properties (S, E, H, D, and A), W_{ik} is the actual value of physicochemical property k of atom i , and $W_{\text{probe},k}$ is the value of the probe atom. α is the attenuation factor and the default value of 0.3 was used. A Gaussian type distance dependence was used between the grid point q and each atom i of the molecule.

Fig. 1 Representative skeletons and the most active molecules showing different regions which are used in contour analysis. **(a)** Compound 41 in the group of 1,4-benzodiazepine-2,5-diones analogs. **(b)** Compound 102 in the class of N-acylpolyamine analogs

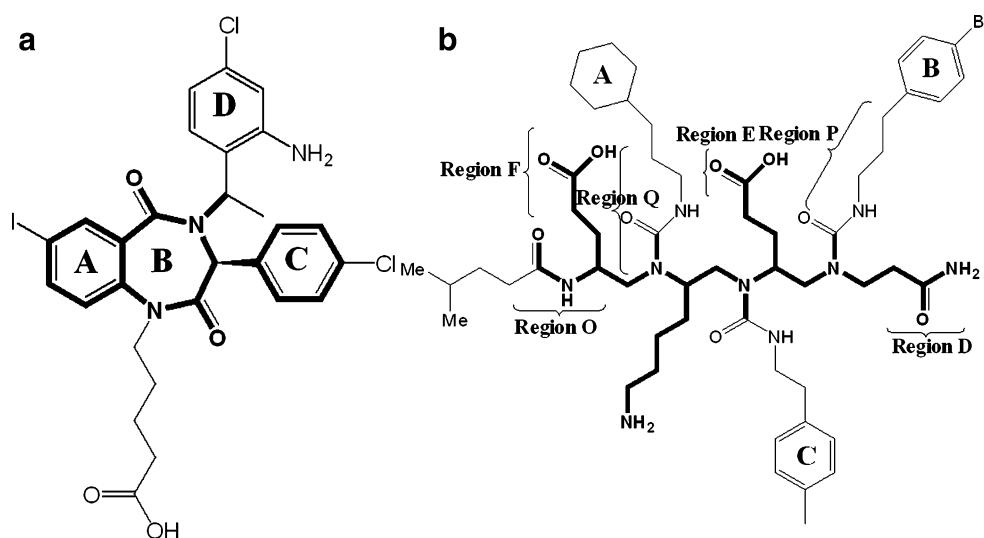
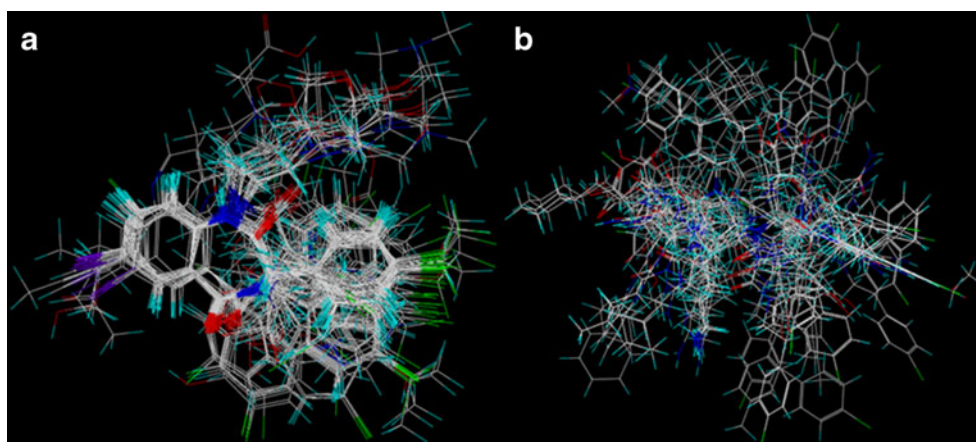


Fig. 2 Ligand-based alignment.

(a) Alignment of 1,4-benzodiazepine-2,5-diones derivatives with compound 41 selected as the template. (b) Alignment of N-acylpolyamine derivatives with compound 102 selected as a template



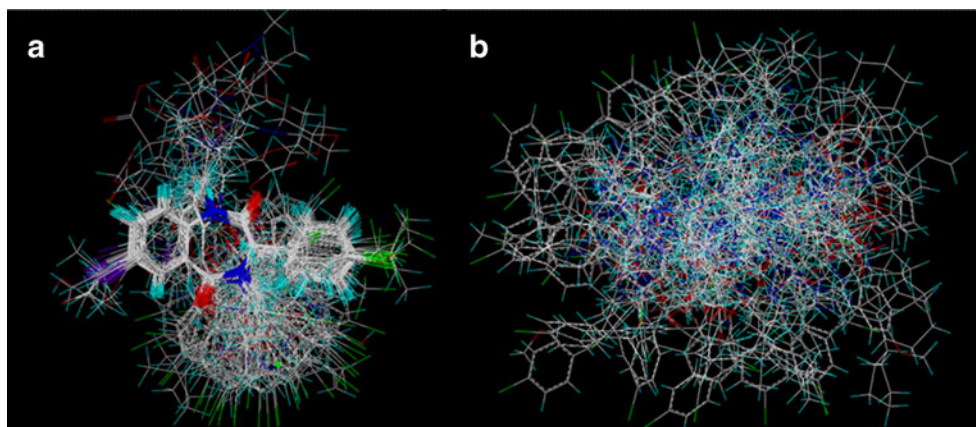
Partial least-square (PLS) method was often used for all 3D-QSAR analyses [36]. It was used to linearly correlate the CoMFA and CoMSIA descriptors to the inhibitory activity values. Cross-validation based on the training set was performed by the leave-one-out (LOO) procedure to determine the optimum number of components (ONC) and the coefficient R^2_{cv} . Then the ONC was further employed to do non-validation analysis to produce non-cross-validated correlation coefficient R^2_{ncv} . The molecules which were not included in the training set were used to evaluate the predictive ability of the models derived using the training set. The predictive correlation coefficient r_{pred}^2 was calculated using Eq. 2:

$$r_{pred}^2 = (SD - PRESS) / SD, \quad (2)$$

where SD is the sum of the squared deviation between the actual pIC_{50} values of the compounds in the test set and the mean pIC_{50} value of the training set ones. PRESS is the sum of squared deviation between predicted and actual activities of the test set compounds.

Fig. 3 Receptor-guided consensus

dynamics alignment. (a) Alignment of 1,4-benzodiazepine-2,5-diones derivatives with compound 41 selected as a template. (b) Alignment of N-acylpolyamine derivatives with compound 102 selected as a template



Molecular docking

To investigate the ligand-protein interactions, all the molecules were docked into the binding site of the receptor using the Surflex module in SYBYL package. During the docking procedure, standard parameters were employed. Crystal structures of HDM₂ (1T4E, 1RV1) were retrieved from RCSB Protein Data Bank (<http://www.pdb.org>). Before docking, all water molecules were removed from the X-ray structure, which is based on two reasons. 1) Normally, it is a standard procedure to remove waters from X-ray structures before docking, because in most cases, the studied compounds are different from the ligands in the crystal structure. Therefore, the interactions between the original ligands and waters would not be true for other different compounds. This phenomenon is also found in this work. For 1T4E, in the crystal structure, only one water molecule forms H-bond with the -NH of ring B for compound 19. However, this H-bonding interaction will disappear for compound 41 as the hydrogen atom of -NH was substituted by -(CH₂)₄COOH. 2) The MD simulations have been also performed to investigate if there are possible

effects provided by water molecules. The MD results revealed that no waters were found H-bonded with the ligands as those in both 1RV1 and 1T4E structures, which further demonstrated that the deletion of waters before docking is reasonable. Subsequently, the relative ligands were extracted and polar hydrogen atoms were added. Protomol, served as an idealized representation of a ligand that made every potential interaction with the binding site, was used to guide molecular docking. During docking process, two parameters, i.e., *protomol_bloat* and *protomol_threshold*, were applied to determine how far from a potential ligand the site should extend and how deep into the protein, the atomic probes were used to define the protomol could penetrate. Ligand docking method was employed presently to generate the protomol, and finally each conformer of all inhibitors in two classes was docked into the binding site 10 times with the *D_score* [37], *G_score* [38], *Chemscore* [39] and *PMF_score* [40] values further used to evaluate the docking analysis. Finally, the top ranked conformations for each molecule were extracted, then aligned together and subsequently utilized in CoMFA and CoMSIA modeling.

MD simulations

The MD simulations were performed with GROMACS software package [41] using the GROMOS96 force field [42]. The molecular topology files for compounds BDPs and NAPA were generated by the program PRODRG 2.5 [43–46], which are shown in the [supporting information](#). For MD simulation, a cubic periodic box of side length of 67.44 Å and 64.70 Å was applied for each class of the inhibitors. The minimum distance between the protein and box walls was set to larger than 8 Å. Three Cl⁻s were employed to neutralize the net charge of each system, i.e., 1T4E.pdb and 1RV1.pdb, respectively. The total number of the atoms was 28057 for 1T4E and 24574 for 1RV1 including the protein complexes and waters, respectively. The remaining box volume was filled using the extended simple point charge (SPCE) water [47].

Prior to the simulation, an energy minimization was applied to the full system without constraints using the steepest descent integrator for 5000 steps, then the system was equilibrated via a 200 ps MD simulations at 300 K. Finally, a 5 ns simulation was performed with a time step of 2 fs. During MD simulation, the standard parameters and main calculation methods were set as follows:

Both models (BDPs, NAPA) used NPT ensemble at 300 K with periodic boundary conditions, the temperature was kept constant by the Berendsen thermostat, the values of the isothermal compressibility were set to $4.5 \times 10^{-5} \text{ bar}^{-1}$ while the pressure was maintained at 1 bar using the Parrinello-Rahman scheme [48], electrostatic interactions were calculated using the particle mesh Ewald method [49],

cut-off distances for the calculation of Coulomb and van der Waals interactions were 1.0 and 1.4 nm, respectively. All the MD simulations lasted 5 ns to ensure that the whole systems were stable.

Results and discussion

3D-QSAR model

Table 1 summarizes the optimum results derived from the CoMFA models where the predictive ability of the models was assessed by predicting the inhibitory ability of the test set molecules. The developed 3D-QSAR models were analyzed by a number of parameters including the cross-validated correlation coefficient (r^2_{cv}), standard error estimate (SEE) and F-statistic values (F), etc. For CoMFA and CoMSIA analysis, two alignment methods were applied to produce the models, with results that the models derived from receptor-guided consensus dynamics alignment were much worse than those derived from ligand-based ones. Consequently, we mainly focused on the models derived by ligand-based method. S and E fields were generated for CoMFA models, while H, D and A, three other fields in addition to the S and E were generated for CoMSIA ones. Since it is argued that the five different descriptor fields may not be totally independent of each other and such dependencies of individual fields usually decrease the statistical significance of the models [50, 51], in the present work all 31 possible combinations of the descriptors for each group were attempted to build the optimum models with highest R^2_{cv} values and other proper statistical results for each class. Figure 4 illustrates the observed versus predicted activities of the optimal CoMFA models for the two classes of dataset. As to the results of other combinations of CoMFA or CoMSIA descriptors, Tables S3 to S6 in the supporting information give a full summation.

BDPs

For this class of inhibitors, the data set was divided into a training set of 59 and a test set of 17 molecules with the statistical parameters obtained from the CoMFA model summarized in Table 1. The best predictions were obtained by the CoMFA standard model using five components, with statistical results of $R^2_{cv}=0.41$, $R^2_{ncv}=0.855$, F value =62.39 and a low SEE=0.26 for the training set obtained, which suggested that this model should be considerably reliable to predict the IC₅₀ values. The steric and electrostatic contributions were 38.7% and 61.3%, respectively, which indicated that the electrostatic field made more contribution to ligand binding affinity. Overall, this model showed reasonable statistical features. However, the CoMSIA model derived

Table 1 The optimum 3D-QSAR results for the two classes of HDM₂ inhibitors

Parameters	1,4-benzodiazepine-2,5-diones				N-acylpolyamine			
	Ligand-based		Structure-based		Ligand-based		Structure-based	
R^2_{cv}	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
SEE	0.405	0.506	0.101	0.06	0.414	0.474	0.07	0.12
F	0.259	0.244	0.38	0.096	0.184	0.097	0.087	0.059
R^2_{pred}	62.388	51.394	56.289	257.36	34.862	75.878	413.34	919.321
SEP	0.6017	0.314	0.033	0.424	0.6913	0.5941	0.116	0.17
Nc	0.523	0.486	0.626	0.692	0.473	0.495	0.549	0.535
Field contribution	5	7	2	10	5	9	2	2
S	0.387	-	0.345	0.143	0.461	-	0.386	0.389
E	0.613	-	0.655	0.556	0.539	0.395	0.614	-
H	-	0.529	-	-	-	0.605	-	-
D	-	0.471	-	-	-	-	-	0.611
A	-	-	-	0.301	-	-	-	-

R^2_{cv} =Cross-validated correlation coefficient using leave-one-out method; SEE=Standard error of estimate; F=Ratio of R^2_{ncv} explained to unexplained= $R^2_{ncv}/(1-R^2_{ncv})$; R^2_{pred} =Predicted correlation coefficient for the test set of compounds; SEP=Standard error of prediction; Nc=Optimal number of principal components; S=steric, E=electrostatic, H=hydrophobic, D=H-bond donor, A=H-bond acceptor.

from the same training set showed poor internal predictions (R^2_{cv} =0.13) using the same S and E fields. Meanwhile, incorporation of H, D and A fields also could not improve the model performance. All these results suggested that the CoMFA model was superior to the CoMSIA one for this set of compounds.

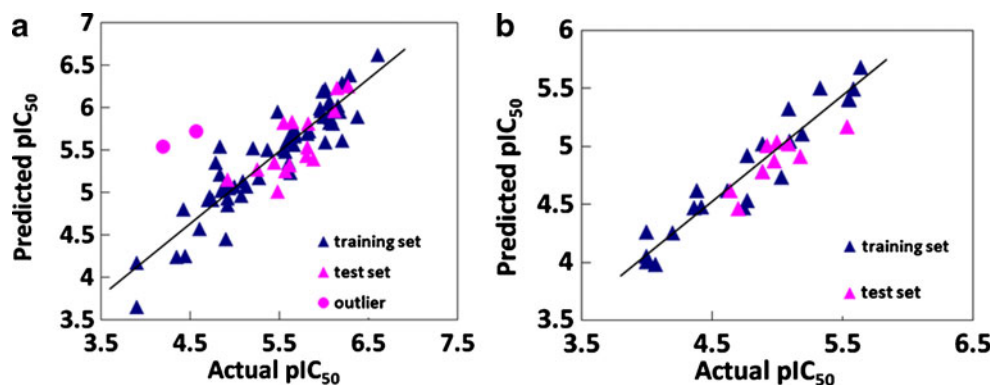
These models were further validated using the external test compounds. The optimum CoMFA model gave good predictions exhibiting the large R^2_{pred} value of 0.60. However, the CoMSIA models possessed poor predictions with low R^2_{pred} value. The correlation between the predicted activities and the actual activities are displayed in Fig. 4a. From this figure, we can see that all points are uniformly distributed along the regression line which suggests that no systematic error exists in the method. Additionally, in the optimal CoMFA model, compounds 10 and 61 were regarded as outliers, since their prediction errors were larger than 1.2 standard deviation (SD). We also

found that for compound 10, which possessed a low inhibitory activity and was the only molecule had a cyclohexyl substituent at the R₂ position. As for compound 61, its outlier status might be due to the low inhibitory activity, which was beyond the forecast range.

NAPA

Using both S and E fields with five components, a best CoMFA model for NAPA inhibitors was obtained, giving a good internal predictive ability illustrated by an R^2_{cv} of 0.41, R^2_{ncv} of 0.91, F value of 34.86 and a low SEE value of 0.18. The steric and electrostatic contributions were found to be 46.1% and 53.9%, respectively. Therefore, the electrostatic field had a greater effect than the steric field on the inhibitory activity, indicating that the electrostatic interactions of the molecules with the receptor could be an important factor for HDM₂ antagonistic activity. However, the CoMSIA model

Fig. 4 Graphs of the predicted pIC_{50} versus the experimental pIC_{50} values of the optimal CoMFA models. (a) CoMFA model of 1,4-benzodiazepine-2,5-diones. (b) CoMFA model of N-acylpolyamine derivatives



using the same fields (SE) exhibited a low R^2_{cv} of 0.34, R^2_{ncv} of 0.84, F value of 17.32 and SEE value of 0.25. Other combinations like SHE, SEHD, SEHA, SEDA, SEHDA were also performed to build models, ending in an average $R^2_{cv} > 0.4$ and $R^2_{pred} > 0.6$ (shown in Table 1 and S4). Overall, the performance of the CoMFA model was superior to that of the CoMSIA one.

To test the predictive ability of the model, a test set of nine molecules excluded from the model derivation was employed. The predictive correlation coefficient R^2_{pred} was 0.69. The graph of actual activity versus predicted pIC_{50} of the training set and test set was illustrated in Fig. 4B, the plots represent a uniform distribution around the regression line with respective slope and intercept very close to one and zero, indicating the satisfactory predictive capability and accuracy of the model.

3D-QSAR contour maps

To visualize the results of the CoMFA and CoMSIA models, contour maps which denote the areas where the molecules would favorably or unfavorably interact with the receptor were generated. For steric field, the green contours represent areas where bulky group would enhance activity while the yellow contours represent areas where bulky group would decrease the activity. In the electrostatic field, the red contour represents the electronegative charge favorable area while the blue color represents the electropositive charge favorable area.

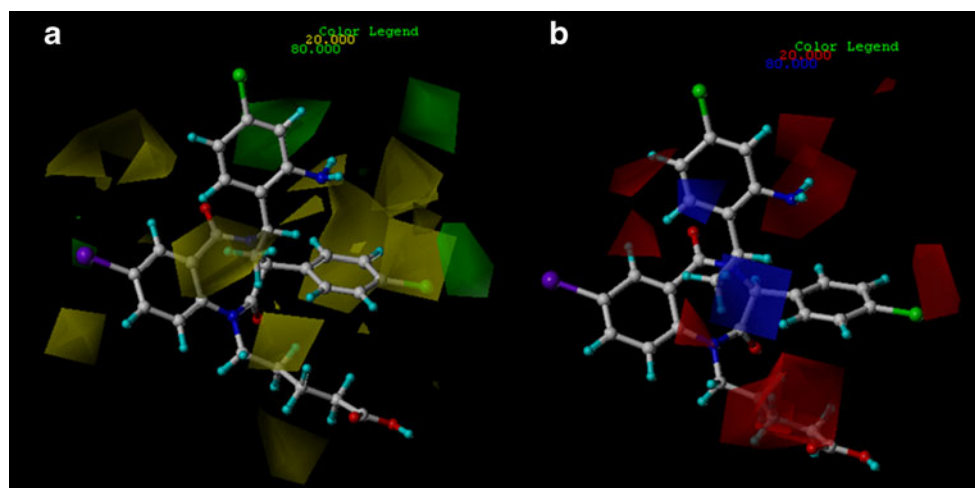
BDPs

CoMFA steric and electrostatic contributions for this class of inhibitors with the most potent compound 41 as a reference were shown in Fig. 5. A small green contour at position 7 of ring A indicates that a sterically bulky group was favored in this region, which is coincide with compounds 29 and 30. Interestingly, compared with compound 33, compound 34

has a larger substituent at this position, leading to an increased inhibitory activity. Sterically unfavored yellow regions were found around ring C suggesting the need for small substituent at this region to enhance the inhibitory activity, for example, the activity of compound 19 (–Cl) was shown to be higher than that of compound 20 (–Br). There were two large green contours observed outside the ring C and around the yellow contours respectively, which indicates that if the substituents at the para position of ring C could touch these green contours, large groups would enhance the activity, which may exactly be the reason why compounds 10, 11, 12, 13 with small groups were less potent than molecules 14, 15, 16 with bulky substituents at the same location. Another yellow contour located between ring B and C, suggesting the desire of the position for groups with small steric barrier to increase the activity, such as compounds 57, 58, 59 and 60. Two yellow contours were also observed near R_3 substituent, indicating the requirement of less bulky substituents in this region for more potent p53-HDM₂ inhibitors.

Figure 5b shows the CoMFA electrostatic field. A medium sized red isopleth at the para position of ring C represents an area where negative charge is favorable, for example, though all having a common structure, compounds 1 to 7 are slightly different at the para position of ring C, where molecules 1, 6, 7 possess negatively charged groups (–CF₃, –Cl, –OCF₃ respectively) and thus increased activity, while chemicals 2, 3, 4, 5 having positive charged substituents (–H, –CH₃, –CH₂CH₃, –CH(CH₃)₂ respectively) at the position and decreased activity. A large red contour around the negatively charged carboxyl group at the R_3 substituent indicates that the negatively charged carboxyl or other negative groups at this position were necessary to increase the activity. Another one big and two medium sized red contours at the -ortho, -meta position of ring D suggests that electronegative substituents would increase the activity, bringing us a conclusion that modifications such as incorporation of electronegative substituents in these areas may

Fig. 5 CoMFA StDev*Coeff contour plots for 1, 4-benzodiazepine-2,5-diones analogs in combination of compound 41. **(a)** The steric contour map, where the green and yellow contours represent 80% and 20% level contributions, respectively. **(b)** The electrostatic contour map, where the blue and red contours represent 80% and 20% level contributions, respectively



greatly improve the inhibitory activity. Meanwhile, a red and a blue contour simultaneously appeared at the linker area between rings B and D, i.e., the electropositive and electro-negative contours emerged at the same region showing that a balance of these properties among the groups presented at this region is required for optimum binding.

NAPA

The CoMFA steric map (Fig. 6a) displays a medium sized green contour at the para position of ring B suggesting that bulky groups at this position would significantly improve the affinity. Thus, compound 102, which has –Br located at the green region, showed more potent inhibitory activity than those compounds of 101, 103 and 104 which possess rather smaller substituent (F) at the same location. Another large green contour around ring A suggests that steric bulky substituents are favored there. A large region of yellow contour around Region P represents the disfavor of bulky group at the region, which is consistent with compounds 103 and 104.

The electrostatic contour map is displayed in Fig. 6b. Clearly, an electropositive favored contour was observed outside the para position of ring B, which is well illustrated by the decreased activity of compound 87 with an oxygen atom touching the blue contour. It is also validated by compound 85 which also exhibited lower activity than compound 102 by extending a substituent of –Cl to the blue contour. One big red contour in the vicinity of region O suggests that electronegative substituent could favor the activity, the fact was that all of the derivatives involved in the present study possessed electronegative substituents (–NH) at this site.

Molecular dynamics simulations and comparison with 3D contour maps

Molecular dynamic simulations were performed to obtain the ‘real’ bioactive conformation which could not be

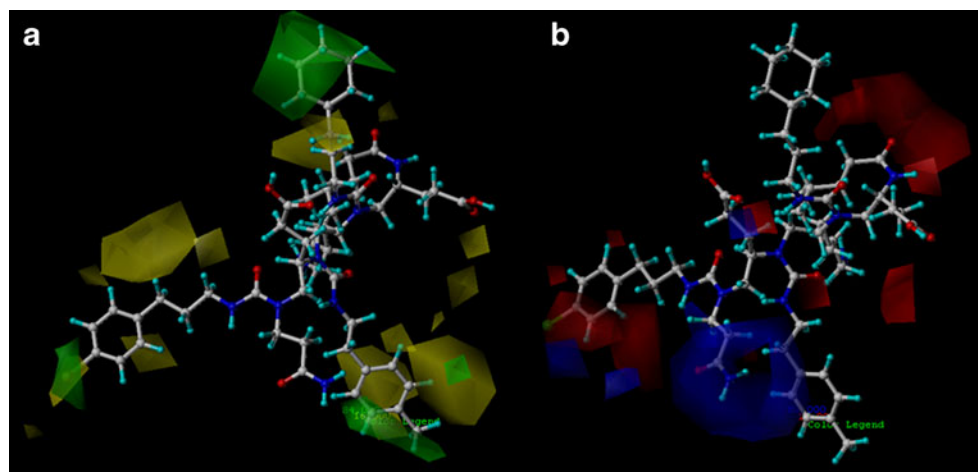
acquired only from the crystal structure of protein–ligand complex [48]. During molecular dynamics simulation, the protein complex was placed in a bulk system which can mimic the real physiological conditions, which can fully consider the flexibility of the receptor, thus making the complex structure more reasonable. However, in the general molecular docking procedure, including the Surflex program, the receptor is normally treated rigidly to model the ligand-receptor binding mode. Therefore, in this work, in order to obtain the optimal conformation for complexes, the docked complex structures of 1T4E with compound 41, 1RV1 with 102 were treated as the initial structures for molecular dynamics simulations. The root-mean-square deviation (RMSDs) of the trajectory with respect to their initial structures ranged from 1.3 to 2.5 Å, 1.3 to 1.8 Å for 1T4E and 1RV1, respectively (Fig. 7a and Fig. 8a). The small RMSDs variations indicate the system of the complexes reached about 2.2 Å, 1.5 Å after 2 ns, and retained this value throughout the simulation.

The superposition of the average structures of the whole trajectory and the docked structures was shown in Fig. 7b and 8b. The comparisons of binding models from MD simulation and the QSAR maps are shown as follows:

BDPs

Figure 9b indicates that the main amino acid residues in the active site cavity responsible for significant interactions were Gln18, Gln24, Lys51, Gly16, Ser17, Leu54, Leu57, Val93, Phe55 and His73. The most potent compound 41 formed three H-bonds: 1) the keto oxygen of ring B (H-bond acceptor) hydrogen-bonded with the –NH group of Ser17 (–O⋯HN, 1.73 Å, 121.1°); 2) the carboxyl oxygen atom at R₃ substituent formed two H-bonds with the –NH₂ of Gln24 (–O⋯HN, 1.91 Å, 152°) and the –NH₂ of Gly16 (–O⋯HN, 2.58 Å, 140°). In addition, ring D of compound 41 formed π - π interaction with His73. As illustrated in Fig. 9b, the R₃ substituent was surrounded by electropositive Lys51 and

Fig. 6 CoMFA StDev*Coeff contour plots for N-acylpolyamine analogs in combination of compound 102. **(a)** The steric contour map, where the green and yellow contours represent 84% and 16% level contributions, respectively. **(b)** The electrostatic contour map, where the blue and red contours represent 82% and 18% level contributions, respectively



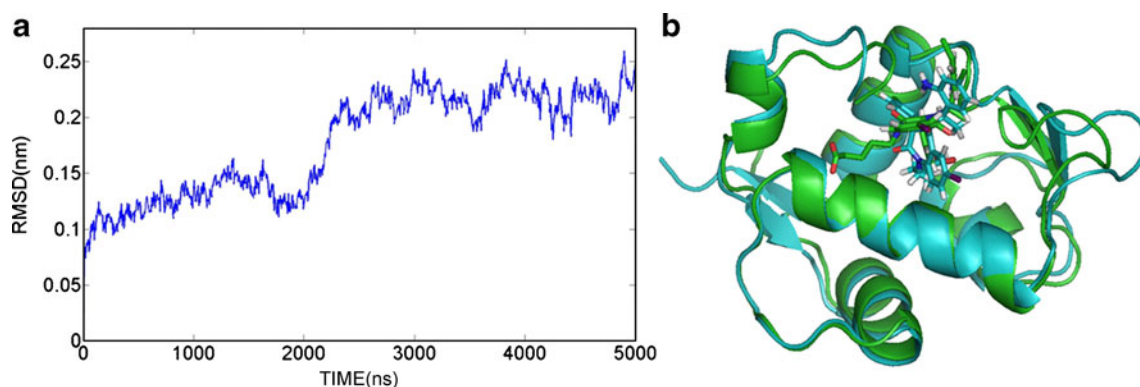


Fig. 7 (a) Plot of the root-mean-square deviation (RMSD) of docked complex versus the MD simulation time in the MD-simulated structures. (b) View of superimposed backbone atoms of the lowest energy structure of the MD simulation (green) and the initial structure

(cyan) for compound 41-HDM₂ complex. Compound 41 is represented as carbon-chain in cyan for the initial complex and carbon-chain in green for the lowest energy complex

Gln24 indicating that electronegative groups at this position are favorable for the inhibitory activity, which result matches well with the electrostatic contour map of the CoMFA model (depicted in Fig. 5b). In addition, the MD-simulated structures of the binding site shows that residues His73, Leu54, Val93 appeared near R₁ substituent, where electronegative substituents would favor electrostatic interaction between ligands and HDM₂. This result is well consistent the electrostatic contour map of the CoMFA model (Fig. 5b), which owns a red contour at the same location. It is worthy to note that near ring C there is sufficient room to accommodate a large hydrophobic substituent into the hydrophobic pocket formed by Leu54, Leu57, Gly58, Gln72 and Val93. This observation is in full agreement with the steric contour map with a green contour (Fig. 5a) around this region. In addition, several yellow contours were observed at this position suggesting that too large substituents would make clash effects on the residues.

Additionally, R₁ substituent was surrounded by some non-polar amino acids Leu54, Leu57, Ile61 and Ile99,

suggesting that non-polar groups at this position is favorable for the binding affinity, this is consistent with all compounds employed in this work possessing a non-polar benzene ring in this region. In addition, the R₃ substituent is surrounded by polar residues Gln24 and Lys36, therefore, a relatively polar substituent at this position is needed. This is in agreement with the fact that the binding affinities of compound 41 with polar substituent –COOH is more active than compounds 43, 44 and 45 which possess non-polar –CH₃ groups in this position.

NAPA

Figure 8b shows the conformation derived from compound 102 with the allosteric binding site of 1RV1, where compound 102 was suitably localized at the binding site. The MD-based docking simulation showed that Glu25, Thr26, Tyr100, Tyr104, Met50, Lys94, His96 and Arg97 were the main residues presented at the binding site and main contributors to the ligand-receptor interaction (shown

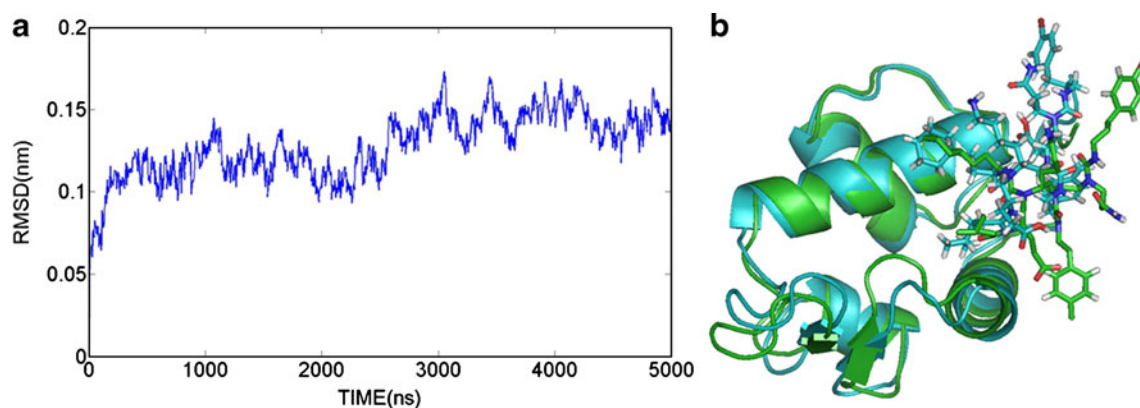
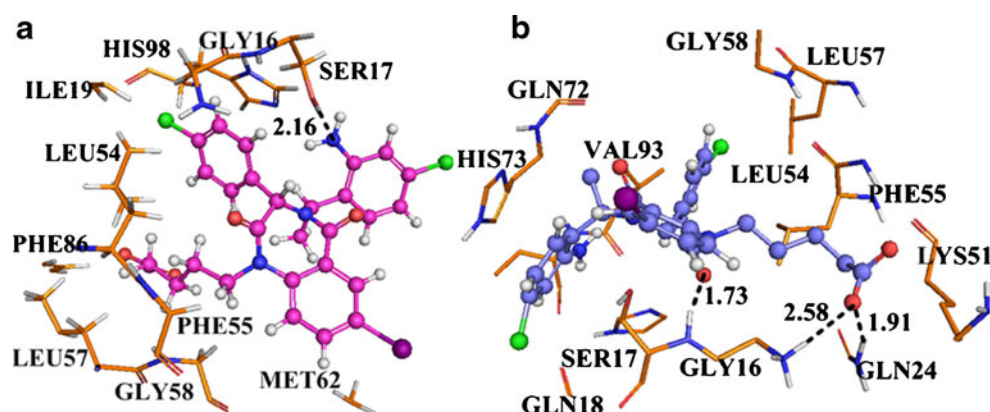


Fig. 8 (a) Plot of the root-mean-square deviation (RMSD) of docked complex versus the MD simulation time in the MD-simulated structures. (b) View of superimposed backbone atoms of the lowest energy structure of the MD simulation (green) and the initial structure

(cyan) for compound 102-HDM₂ complex. Compound 102 is represented as carbon-chain in cyan for the initial complex and carbon-chain in green for the lowest energy complex

Fig. 9 (a) Docked conformation derived for compound 41 in complex to the active site of HDM₂ protein. (b) Plot of the MD-simulated structures of the binding site with ligand 41. H-bonds are shown as dotted black lines, active site amino acid residues are represented as sticks, the inhibitors are shown as stick and ball model



in Fig. 10b). And a total of nine hydrogen bonds were formed in this system. The carboxyl oxygen at region Q acted as an acceptor to form H-bond with the side chain hydroxyl group of Tyr100 ($-O\cdots HO$, 1.42 Å, 175.6°), the carboxyl oxygen at region F formed two H-bonds with the side chain NH of Arg97 ($-O\cdots HN$, 1.95 Å, 139.6°) and ($-O\cdots HN$, 1.76 Å, 153.8°). In addition, the carboxyl oxygen at region E interacted through H-bonding with the backbone NH of Glu25 ($-O\cdots HN$, 1.81 Å, 171.8°) and the OH of Thr26 ($-O\cdots HO$, 1.52 Å, 161.9°), also the hydroxyl oxygen atom formed two H-bonds with Glu25 ($-O\cdots HN$, 2.36 Å, 114.9°), ($-O\cdots HN$, 2.74 Å, 90.7°). The carboxyl oxygen at region P was involved in hydrogen bonding with the backbone NH₂ of Glu25 ($-O\cdots HN$, 2.57 Å, 89.4°), ($-O\cdots HN$, 2.64 Å, 85.6°). As displayed in Fig. 10b, several basic electropositive residues Arg97, His96 were close to region O which suggests that electronegative substituents at these positions would enhance the inhibitory activities. This could be confirmed by the electrostatic contour map (Fig. 6b) which holds a red contour at this position. A large binding cavity formed by residues Lys94, His96, Arg97 and Thr100 over ring A could contain a bulky substituent, which is corroborated by the contour analysis results of the green contour obtained from the CoMFA model (Fig. 6a). Meanwhile, ring C established arene-cation

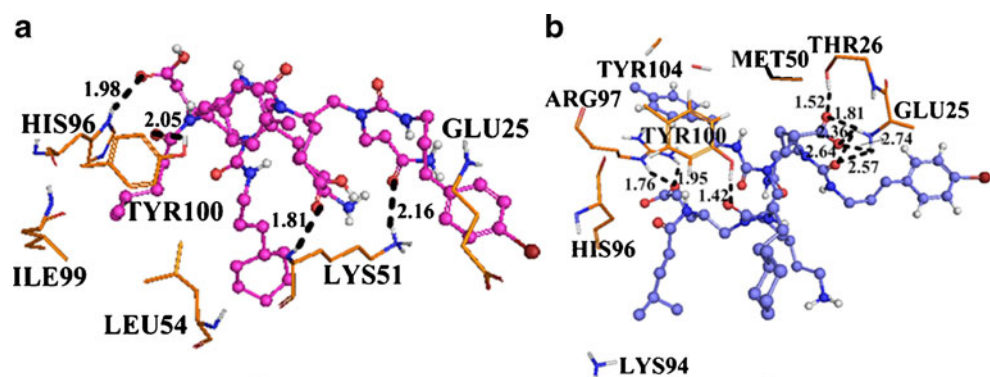
interaction with Arg97 which further enhanced the stability of the ligand-receptor interaction.

Moreover, the polar substituent $-COOH$ of region E extends to polar amino acids residues Glu25 and Thr26, indicating that polar groups are favorable for the inhibitory activity. This is the same as the $-COOH$ of region F which is surrounded by Arg97 and His96. It is remarkable to note that the non-polar amino acid residues Phe55 and Leu54 are near ring A, which indicates that the non-polarity of ring A is needed for the inhibitory activity.

Comparison of the results of docking and MD simulation

When predicting the binding mode of the ligand to receptor, molecular docking provides a good starting to evaluate the stability of the predicted interactions involved in binding [51]. While molecular dynamics was employed for further calculations considering the flexibility of the receptor and the effect of water solvation. In this study, firstly the most potent compounds (41 and 102) were docked into the binding site of the receptors, which complexes were further used to undertake MD simulations. As depicted in Fig. 7b and Fig. 8b, the initial conformations changed a lot after the MD simulations. Furthermore, we further examined the stability of the hydrogen bond formed during the docking process by calculating the presence rates of hydrogen bonds

Fig. 10 (a) Docked conformation derived for compound 102 in complex to the active site of HDM₂ protein. (b) Plot of the MD-simulated structures of the binding site with ligand 102. H-bonds are shown as dotted black lines, active site amino acid residues are represented as sticks, the inhibitors are shown as stick and ball model



during MD simulation. The hydrogen bonds produced between all possible donors D and acceptors A. A geometrical criterion shown in Eq. 3 is used to determine if the hydrogen-bond exist.

$$\begin{aligned} r &\leq r_{\text{HB}} = 0.35 \text{ nm} \\ \alpha &\leq \alpha_{\text{HB}} = 30^\circ \end{aligned} \quad (3)$$

As seen from Fig. 9 and Fig. 10, it is concluded that the presence of corresponding H-bonds between the ligands and receptors exhibited higher frequencies than the docking analysis.

For BDPs analogs, one hydrogen bond was formed based on docking simulation between the amino nitrogen atom in ring D and the hydroxyl group of Ser17 ($-\text{N}\cdots\text{HO}$, 2.16 Å, 177.1°), which was yet not preserved during MD simulation. It was interesting to note that three new hydrogen bonds were constructed to enhance the stability of the complex.

As for the lost hydrogen bond, it could be compensated by the electrostatic interactions. As illustrated in Fig. 9b, the electronegative nitrogen atom was surrounded by electropositive residues His73 and Gln18 which could still be involved in electrostatic interactions.

During docking simulation, four hydrogen bonds were formed in compound 102 and 1R1V1 (Fig. 10a), i.e., region O oxygen atom...OH of Tyr100 ($-\text{O}\cdots\text{HO}$, 2.05 Å, 108.8°), region F oxygen atom...NH of His96 ($-\text{O}\cdots\text{HN}$, 1.98 Å, 158.2°), region E oxygen atom...NH of Lys51 ($-\text{O}\cdots\text{HN}$, 1.81 Å, 143.8°), region D oxygen atom...NH of Lys51 ($-\text{O}\cdots\text{HN}$, 2.16 Å, 147.7°). However, after MD simulations, only two hydrogen bonds were preserved, and the amino acids involved in forming the H-bonds were also changed as explained above. In a word, comparatively the molecular dynamics simulation provides a more stable conformation, so we chose the MD-simulated structures for further 3D-QSAR studies.

Comparison of binding modes for each class

In order to explore the differences and similarities between the two classes of p53-HDM₂ inhibitors, we compared the binding mode of each class. The CoMFA steric field in the BDPs model makes more contributions to the binding activity, as well as the NAPA derivatives. In addition, the docked models reveal that the hydrogen bonding is also important for interactions between the ligand and the target. For BDPs, three H-bonds were formed between the most potent compound 41 and Ser17 ($-\text{O}\cdots\text{HN}$, 1.73 Å, 121.1°), Gln24 ($-\text{O}\cdots\text{HN}$, 1.91 Å, 152°), Gly16 ($-\text{O}\cdots\text{HN}$, 2.58 Å, 140°), at the same time, compound 41 also formed π - π interaction with His73. The NAPA derivative 102 formed a total of nine hydrogen bonds, mainly through interacting with Tyr100, Arg97, Glu25 and Thr26. Additionally, arene-cation interaction has been established between Arg97 and

this compound. The NAPA derivatives formed more hydrogen bonds than the BDPs analogs which also demonstrate why they were more active than BDPs. In addition, from the comparison of interacting patterns for the two series of inhibitors, we also found that there were no common amino acids which were involved in forming H-bonds among these two classes. Thus a conclusion can be made that these inhibitors might have different binding modes with p53-HDM₂.

Conclusions

In this study, 3D-QSAR models for these two different classes of p53-HDM₂ inhibitors, i.e., BDPs and NAPA analogs were built using both the ligand-based and receptor-guided alignments to find out the structural relationship with the activity. We also compared two alignment schemes, with respect to the internal predictive ability and the robustness of their respective models for structurally diverse data sets. As a result, using the ligand-based alignment was superior to that based on the receptor alignment and the CoMFA models were superior to the CoMSIA ones. In addition to the structural features of steric and electrostatic fields, this article still elucidated the structural features of hydrogen bond related to the HDM₂ activity. The external test validation results indicate that the CoMFA models could be successfully used for predicting the inhibitory efficiency of these two classes of compounds. The optimal models from the BDPs and NAPA analogs implied the significant roles of electrostatic field effects on their binding affinity. For the BDP model, the bulky substituent at ring C played a main contribution to inhibitory activity. The crucial role of R₃ group was revealed by the fact that it formed strong hydrogen bond with its receptor and the more potent compounds had bulky groups at the para position of ring A, small and electronegative groups around ring D. In addition, the non-polarity of R₁ and the polarity of R₃ had important effect on the binding affinity. For the NAPA model, the electropositive and bulky substituents at ring B had a major impact on the inhibitory activity, and hydrogen-bond interactions also improve the compound potency. For the more active compounds, they possessed bulky and electropositive groups at ring B, large and non-polar groups at ring A and electronegative substituents in region O.

In addition, molecular dynamics simulations were also performed for the docking complexes. The analysis revealed that the steric, electrostatic, and H-bonding interactions between ligands and key amino-acid residues in binding pocket of HDM₂ correlated well with CoMFA analysis and data of the inhibitory activity. The binding modes showed that for BDPs, the key residues were Gln24,

Lys51, Gly16, Ser17, Leu54, Gln72, Gly58, Leu57, Val93, Phe55 and His73. Yet in the NAPA-protein complex, those key residues changed to Glu25, Thr26, Tyr100, Tyr104, His96, Leu54, Met50 and Phe55. Further analysis showed that the CoMFA contour maps had a good correlation with the docking analysis after MD simulations, further proving the reliability of the models and promoting the understanding of the ligand-receptor interactions.

In this study, the CoMFA analyses gained some insights into the key structural factors affecting the bioactivity of these inhibitors. The excellent predictive ability of the developed CoMFA models indicated that they could be used for predicting the IC₅₀ values of this two class of inhibitors. Furthermore, the CoMFA contour maps along with the docking results after MD simulation offered enough information to understand the structure–activity relationship and identify structural features influencing the inhibitory activity. Overall, the correlation of the results obtained from 3D-QSAR, molecular dynamics simulation and molecular docking studies could, we hope, be served as a useful guideline for further modification and designing of new compounds as p53-HDM₂ inhibitors.

Acknowledgments The research is supported by high-performance computing platform of Northwest A & F University. The research is financially supported by the Fund of Northwest A & F University. The authors are grateful to Prof. L. Yang for access of Sybyl software.

References

- Lane DP (1992) *Nature* 358:15–16
- Gudkov AV, Komarova EA (2007) *Hum Mol Genet* 16:R67–72
- Komarova EA, Gudkov AV (2001) *Biochem Pharmacol* 62:657–667
- Vogelstein B, Lane D, Levine AJ (2000) *Nature* 408:307–310
- Feki A, Irminger-Finger I (2004) *Rev Oncol Hematol* 52:103–116
- Kubbutat MH, Jones SN, Vousden KH (1997) *Nature* 387:299–303
- Kussie PH et al (1996) *Science* 274:948–953
- Wu X, Bayle JH, Olson D, Levine AJ (1993) *Genes Dev* 7:1126–1132
- Barak Y, Juven T, Haffner R (1993) *OrenM. EMBO J* 12:461–468
- Momand J, Zambetti GP, Olson DC, George D, Levine AJ (1992) *Cell* 69:1237–1245
- Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ et al (1996) *Science* 274:948–953
- Michael D, Oren M (2003) *Cancer Biol* 13:49–58
- Chen F, Wang W, Wafik S (2010) *El-Deiry. Biochem Pharmacol* 80:724–730
- Vassilev LT (2004) *Cell Cycle* 3:419–421
- Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z et al. (2004) *Science* 303:844–848
- Brown CJ, Lain S, Verma CS, Fersht AR, Lane DP (2009) *Nat Rev Cancer* 9:862–873
- Ding K, Lu Y, Nikolovska-Coleska Z, Qiu S, Ding Y, Gao W, Stuckey J, Krajewski K, Roller PP, Tomita Y, Parrish DA, Deschamps JR, Wang S (2005) *J Am Chem Soc* 127:10130–10131
- Parks DJ, LaFrance LV, Calvo RR, Milkiewicz KL, Gupta V, Lattanze J, Ramachandren K, Carver TE, Petrella EC, Cummings MD, Maguire D, Grasberger BL, Lu T (2005) *Bioorg Med Chem Lett* 15:765–770
- Lu Y, Nikolovska-Coleska Z, Fang X, Gao W, Shangary S, Qiu S, Qin D, Wang S (2006) *J Med Chem* 49:3759–3762
- Chen L, Yin H, Farooqi B, Sebt S, Hamilton AD, Chen J (2005) *Mol Cancer Ther* 4:1019–1025
- Lu F, Chi SW, Kim DH, Han KH, Kuntz ID, Guy RK (2006) *J Comb Chem* 8:315–325
- Shangary S, Wang S (2008) *Clin Cancer Res* 14:5318–5324
- Koblish HK, Zhao S, Franks CF, Donatelli RR, LaFrance LV, Leonard KA, Gushue JM, Parks DJ, Calvo RR, Milkiewicz KL, Marugan JJ, Cummings Raboisson P, MD Grasberger BL, Lu T, Molloy CJ, Maroney AC (2006) *Mol Cancer Ther* 5:160–169
- Li Y, Wang YH, Yang L, Zhang SW, Liu CH (2006) *J Mol Des* 5:1–12
- Li Y, Wang YH, Yang L, Zhang SW, Liu CH, Yang SL (2005) *J Mol Struct* 733:111–118
- Wang X, Yang W, Xu X, Zhang H, Li Y, Wang Y (2010) *Curr Med Chem* 17:2788–2803
- Wei SP, Ji ZQ, Zhang HX, Zhang JW, Wang YH, Wu WJ (2010) *J Mol Model*. doi:10.1007/s00894-010-0765-x
- Hardcastle IR, Ahmed SU, Atkins H, Farnie G, Golding BT, Griffin RJ, Guyenne S, Hutton C, Källblad P, Kemp SJ, Kitching MS, Newell DR, Norbedo S, Northen JS, Reid RJ, Saravanan K, Willems HM, Lunec J (2006) *J Med Chem* 49:6209–6221
- Stoll R, Renner C, Hansen S, Palme S, Klein C, Belling A, Zeslawski W, Kamionka M, Rehm T, Mühlhahn P, Schumacher R, Hesse F, Kaluza B, Voelter W, Engh RA, Holak TA (2001) *Biochemistry* 40:336
- Parks DJ, LaFrance LV, Calvo RR, Milkiewicz KL, Gupta V (2006) *Bioorg Med Chem Lett* 16:3310–3314
- Leonard K, Marugan JJ, Raboisson P, Calvo R, Gushue JM (2006) *Bioorg Med Chem Lett* 16:3463–3468
- Hayashi R, Wang D, Hara T, Iera J, Durell SR, Appella DH (2009) *Bioorg Med Chem* 17:7884–7893
- Klebe G, Abraham U, Mietzner T (1994) *J Med Chem* 37:4130–4146
- Matthew C, Richard DC III, Van Nicole O (1989) *J Comput Chem* 10:982–1012
- Pirhadi S, Ghasemi JB (2010) *Europ J Med Chem* 1–7
- Wold S, Albano C, Dunn WJ, Edlund U, Esbenson K, Geladi P, Hellberg S, Lindburg W, Sjostrom M (1984) *Multivariate data analysis in chemistry*. In: Kowalski BR (ed) *Chemometrics: Mathematics and Statistics in Chemistry*, vol 138, NATO ASI Series- Reidel. Dordrecht, Holland, pp 17–96
- Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) *J Mol Biol* 161:269–288
- Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) *J Mol Biol* 267:727–748
- Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) *J Comput Aided Mol Des* 11:425–445
- Muegge I, Martin YC (1999) *J Med Chem* 42:791–804
- van der Spoel D, van Buuren AR, Tieleman DP, Berendsen HJC (1996) *J Biomol NMR* 8:229–238
- Lindahl E, Hess B, van der Spoel D (2001) *J Mol Model* 7:306–317
- van Aalten DMF, Bywater R, Findlay JBC, Hendlich M, Hooft RWW, Vriend G (1996) *J Comput Aided Mol Des* 10:255–262
- Barreca ML, Ortuso F, Iraci N (2007) *Biochem Biophys Res Commun* 363:554–560
- Liu R, Li X, Li Y, Jin P, Qin W, Qi J (2009) *Biosens Bioelectron* 25:629–634
- Niu C, Xu Y, Xu Y, Luo X (2005) *J Phys Chem B* 109:23730–23738
- Berendsen HJC, Grigera JR, Straatsma TPJ (1987) *Phys Chem* 91:6269–6271
- Parrinello M, Rahman A (1981) *J Appl Phys* 52:7182–7190
- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) *J Chem Phys* 103:8577–8593
- Bohm M, Stürzebecher J, Klebe G (1999) *J Med Chem* 42:458–477
- Bringmann G, Rummey CJ (2003) *J Chem Inf Comput Sci* 43:304–316

Noscapinoids with anti-cancer activity against human acute lymphoblastic leukemia cells (CEM): a three dimensional chemical space pharmacophore modeling and electronic feature analysis

Pradeep K. Naik · Seneha Santoshi · Harish C. Joshi

Received: 1 January 2011 / Accepted: 22 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract We have identified a new class of microtubule-binding compounds—noscapinoids—that alter microtubule dynamics at stoichiometric concentrations without affecting tubulin polymer mass. Noscapinoids show great promise as chemotherapeutic agents for the treatment of human cancers. To investigate the structural determinants of noscapinoids responsible for anti-cancer activity, we tested 36 structurally diverse noscapinoids in human acute lymphoblastic leukemia cells (CEM). The IC_{50} values of these noscapinoids vary from 1.2 to 56.0 μ M. Pharmacophore models of anti-cancer activity were generated that identify two hydrogen bond acceptors, two aromatic rings, two hydrophobic groups, and one positively charged group as essential structural features. Additionally, an atom-based quantitative structure–activity relationship (QSAR) model was developed that gave a statistically satisfying result ($R^2=0.912$, $Q^2=0.908$, Pearson $R=0.951$) and effectively predicts the anti-cancer activity of training and test set compounds. The pharmacophore model presented here is well supported by

electronic property analysis using density functional theory at B3LYP/3-21*G level. Molecular electrostatic potential, particularly localization of negative potential near oxygen atoms of the dimethoxy isobenzofuranone ring of active compounds, matched the hydrogen bond acceptor feature of the generated pharmacophore. Our results further reveal that all active compounds have smaller lowest unoccupied molecular orbital (LUMO) energies concentrated over the dimethoxy isobenzofuranone ring, azido group, and nitro group, which is indicative of the electron acceptor capacity of the compounds. Results obtained from this study will be useful in the efficient design and development of more active noscapinoids.

Keywords Noscapinoid · Pharmacophore · 3D QSAR · Electrostatic potential profile · Lowest unoccupied molecular orbital · Highest occupied molecular orbital · Free energy of solvation

Introduction

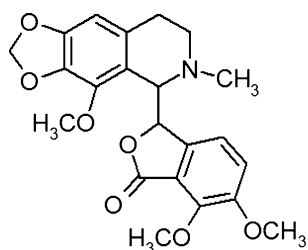
Phthalideisoquinoline [(S)-3-((R)-4-methoxy-6-methyl-5,6,7,8-tetrahydro-[1,3]dioxolo[4,5-g]isoquinolin-5-yl)-6,7-dimethoxyiso-benzofuran-1(3 H)-one], also commonly known as noscapine (Fig. 1), is a plant-derived non-toxic orally available alkaloid that has been used clinically as a cough suppressant in humans and in experimental animals [1–3]. Its role as a stoichiometric tubulin-binding anti-cancer drug was discovered in our laboratory [4]. Noscapine binds to tubulin and alters tubulin conformation as determined by significant changes in the circular dichroism spectra of tubulin [4]. Unlike other clinically used microtubule(MT)-binding drugs such as vincas (MT

P. K. Naik (✉) · H. C. Joshi (✉)
Department of Cell Biology,
Emory University School of Medicine,
615 Michael Street,
Atlanta, GA 30322, USA
e-mail: pnaik2@emory.edu

H. C. Joshi
e-mail: joshi@cellbio.emory.edu

S. Santoshi
Department of Biotechnology and Bioinformatics,
Jaypee University of Information Technology,
Waknaghat, Solan 173215 Himachal Pradesh, India

Fig. 1 Chemical structure of noscapine, which consists of an isoquinoline ring (*top section*) and a dimethoxy isobenzofuranone ring (*lower section*)



depolymerizing) and taxanes (MT over-polymerizing and bundling), noscapine simply suppresses the dynamics of MT assembly without changing the dimer/polymer ratio, blocks mitotic progression, and induce apoptosis in many cancer cell types [4–7]. Our laboratory has now synthesized more potent noscapine derivatives (collectively called noscapinoids) that may be suited for development of yet better anti-cancer agents [8–13]. This array of compounds led us to evaluate quantitatively the structure–activity relationships of noscapinoids.

In this study, we evaluate the structure–activity relationship of noscapinoids by developing pharmacophore models. The quality of pharmacophores was examined by atom-based 3D quantitative structure–activity relationship (QSAR) models [14] using Pharmacophore Alignment and Scoring Engine (PHASE, Schrodinger suite). This engine has been used successfully in many drug discovery studies such as those investigating glycoprotein (GP)IIb/IIIa antagonists [15], H3-antihistaminics [16], MRP1 modulators [17], and dihydrofolate reductase inhibitors [18]. Additionally, we performed several quantum chemical calculations to analyze the role of the calculated electronic properties in identifying a pharmacophore model for anti-cancer activity. Electronic features are particularly helpful, not only in the design of new drugs and in helping to understand their mechanism of action [19–24] but also in predicting potential reactive sites that may lead to off-target effects. Electronic properties such as molecular electrostatic profile (MESP), lowest unoccupied molecular orbital (LUMO), highest occupied molecular orbital (HOMO), and aqueous solvation energy are very useful parameters for an accurate understanding of the chemical reactivity of molecules [19–21, 25]. MESP is a widely used and proven tool for exploring molecular electronic structure, reactivity patterns, and structure–activity relationship studies [22–24].

We believe that our ligand-based approaches to identifying pharmacophore models will provide important insights that will lead to the design of novel noscapinoids as anti-cancer agents. Furthermore, determination of the molecular electronic properties responsible for the potent anti-cancer activity of noscapine and its congeners should pinpoint the fundamental molecular-level forces responsible for their potency.

Materials and methods

Biological data

We used 36 noscapine derivatives (noscapinoids) that were synthesized in our laboratory [8–13] for pharmacophore generation (Table 1). Structural modifications were introduced primarily in the isoquinoline and benzyl furanone rings of noscapine (Fig. 1). Anti-cancer activities of these compounds were measured against CEM, a multi-drug resistant human T-cell lymphoblastoid cell line under identical experimental conditions (to minimize bias). A drug-resistant CEM cell line (overexpressing the drug efflux pumps, MDR and MRP [26]) was generously provided by W.T. Beck (Cancer Center, University of Illinois at Chicago). Cell culture reagents were obtained from Mediatech (<http://cellgro.com>). Cells were grown in RPMI 1640 medium (Mediatech) supplemented with 10% fetal calf serum, 1% antibiotics (penicillin/streptomycin), 2 mM l-glutamine, at 37°C in a humidified atmosphere with 5% CO₂. Cell proliferation assay was performed in 96-well plates as previously described [9, 27]. Briefly, cells were seeded in 100 μL growth medium at a density of 5 × 10³ cells per well in 96-well plates and allowed to establish for 24 h. Serially diluted concentrations (0.01–500 μM) of noscapinoids were then added and cells were incubated for 72 h. We measured the inhibition of cell proliferation in a colorimeter by 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2 H-tetrazolium inner salt (MTS) assay using the CellTiter96 AQueous One Solution Reagent (Promega, Madison, WI). Cells were incubated with MTS for 3 h and absorbance was measured at a wavelength of 490 nm using a microplate reader (Molecular Devices, Sunnyvale, CA). The percentage of cell survival as a function of drug concentration was then plotted to determine the IC₅₀ values (the drug concentration needed to prevent cell proliferation by 50%). The negative logarithm of the measured IC₅₀ value in molar concentration (expressed as pIC₅₀ values) of these compounds was used in the pharmacophore modeling and QSAR study.

Generation of the common pharmacophore hypotheses

In order to rationalize chemical structure with the observed anti-cancer activity, pharmacophore models were developed using a dataset of 36 noscapinoids. Molecular models of these structures were built using the builder panel in Maestro (version 9.1; Schrodinger package, [28]). Each structure was assigned an appropriate bond order using ligprep script [28] and was subsequently subjected to energy minimization using the OPLS 2005 force field with a dielectric constant of 1.0. In the development of common pharmacophore hypotheses (CPHs) for anti-cancer activity

Table 1 Chemical structures of noscapine and its congeners used in the present study, along with their observed inhibitory activity of cell proliferation (CEM cell line)

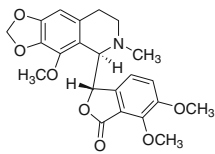
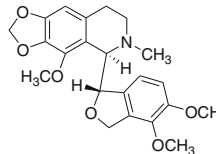
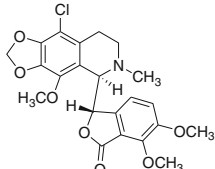
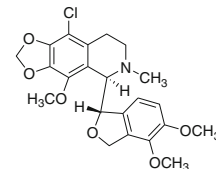
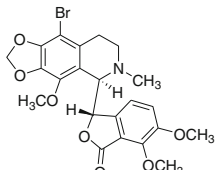
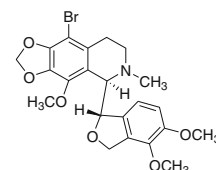
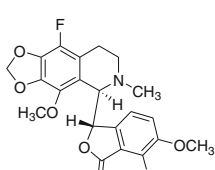
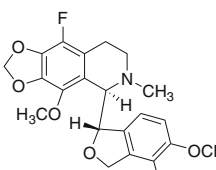
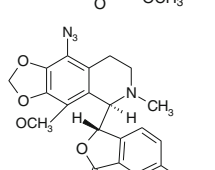
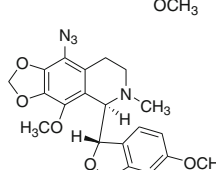
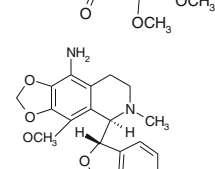
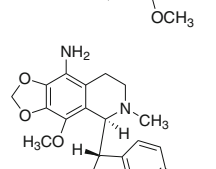
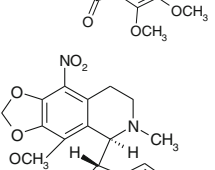
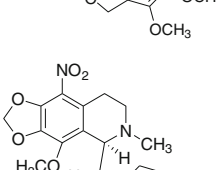
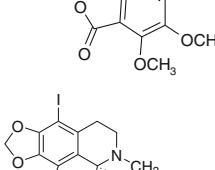
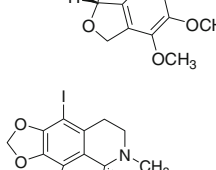
SI No.	Compound Structure	IC ₅₀ (M)	SI No.	Compound Structure	IC ₅₀ (M)
1		16.59 x 10 ⁻⁶	2		28.3 x 10 ⁻⁶
3		1.2 x 10 ⁻⁶	4		45.2 x 10 ⁻⁶
5		1.9 x 10 ⁻⁶	6		2.8 x 10 ⁻⁶
7		2.3 x 10 ⁻⁶	8		15.5 x 10 ⁻⁶
9		2.6 x 10 ⁻⁶	10		2.6 x 10 ⁻⁶
11		3.0 x 10 ⁻⁶	12		3.0 x 10 ⁻⁶
13		10.0 x 10 ⁻⁶	14		10.0 x 10 ⁻⁶
15		38.9 x 10 ⁻⁶	16		30.5 x 10 ⁻⁶

Table 1 (continued)

#	Compound Structure	IC ₅₀ (M)	#	Compound Structure	IC ₅₀ (M)
17		48.0 x 10 ⁻⁶	18		45.2 x 10 ⁻⁶
19		44.2 x 10 ⁻⁶	20		40.6 x 10 ⁻⁶
21		41.9 x 10 ⁻⁶	22		45.1 x 10 ⁻⁶
23		46.5 x 10 ⁻⁶	24		44.4 x 10 ⁻⁶
25		44.1 x 10 ⁻⁶	26		51.8 x 10 ⁻⁶
27		39.5 x 10 ⁻⁶	28		42.6 x 10 ⁻⁶
29		41.0 x 10 ⁻⁶	30		42.3 x 10 ⁻⁶
31		35.5 x 10 ⁻⁶	32		53.6 x 10 ⁻⁶
33		56.0 x 10 ⁻⁶	34		52.8 x 10 ⁻⁶
35		37.4 x 10 ⁻⁶	36		43.3 x 10 ⁻⁶

against the CEM cell line, compounds were divided into active (10 molecules with $pIC_{50} > 5.0$) and inactive (23 compounds with $pIC_{50} < 4.5$) class compounds. PHASE (version 3.2, Schrodinger [28]) was used to generate CPHs. A maximum of 200 conformers were generated for each of

the 36 nescapinoids by sampling the conformational space using a combined Monte-Carlo multiple minimum/low mode (MCMMLMOD) method with OPLS-2005 force field, distance dependent dielectric solvation model, and a maximum of 1,000 iterations of post-minimization. Con-

formers not within 10 kcal mol⁻¹ of the global minimum were discarded. Visual inspection and careful analysis of chemical features were conducted to guide the pharmacophore development process. Pharmacophore variants considered for anti-cancer activity were composed of seven sites with the following chemical features: hydrogen-bond acceptor (A, pink sphere), aromatic ring (R, ring), hydrophobic (H, green sphere), and positively charged group (P, blue sphere). Common pharmacophores were searched for in at least six of the ten active structures with a final box size of 1.0 Å³, a tree depth of four and intersite distances of 2.0 Å. Several hypotheses were generated, of which the best (AAHPPRR) mapped onto compound 3. Based on the PHASE scoring functions [15], this hypothesis was chosen as the final pharmacophore model for use in subsequent screening and QSAR investigations.

Assessment of significant CPH using partial least square analysis

The quality of CPHs was evaluated by developing atom-based 3D QSAR models [15] that correlate chemical structures with anti-cancer activity against the CEM cell line. The data set of 36 molecules was divided randomly into a training set of 26 and a test set of 10 molecules (Table 2) incorporating anti-cancer activity and chemical diversity. Atom-based QSAR models were built for the generated pharmacophores using a grid spacing of 1.0 Å to encompass the space occupied by the aligned training set molecules. Based on the occupation of a cube by a ligand atom of training set molecules, the total number of volume bits was assigned to a given cube. Thus, molecules were represented by a string of zeros and ones, according to the cubes they occupy and the different types of atoms/sites that reside in those cubes. QSAR models were generated by applying partial least squares (PLS) regression to the pool of binary-valued (as independent variables) and using pIC₅₀ (as dependent variable) with three PLS factors. Each of these models was validated using a test set of ten molecules that were not considered during model generation.

Density functional theory calculations

The molecular structures of the noscapinoids obtained from the predictive pharmacophore model were used for density functional theory (DFT) calculations. The structures were energy minimized based on molecular mechanics (Macro-model, version 9.8, Schrodinger package [28]) during pharmacophore model development. However, in order to assess detailed aspects of the electronic structure of the molecules to calculate the various electronic properties accurately, we want the geometry of the molecule to be

optimized at a specific level of theory. Thus, to ensure that the geometry of the structure is fairly reasonable, a complete geometric optimization of the structures is necessary. This is because small changes in the geometry can create very large changes in components of the electronic structure even if the overall self-consistent field (SCF) calculated energy (the total energy) of the structure does not change much. Therefore we carried out complete geometric optimization of the structure by applying hybrid DFT with B3LYP (Becke's three-parameter exchange potential and the Lee-Yang-Parr correlation functional) [29, 30], and using basis set 3-21 G* level [31–33] to calculate the electronic properties of noscapinoids at this level of theory. Similar computational methods have recently been used in calculating the electronic properties of molecules [34, 35]. Various electronic properties: 3D-MESP, dipole moment, LUMO, HOMO, and aqueous solvation energy were calculated using Jaguar (version 7.7, Schrodinger [28]). The MESP $V(r)$ at point r , due to a molecular system with nuclear charges $\{Z_A\}$, located at $\{R_A\}$ and the electron density $\rho(r)$ were derived using the equation:

$$V(r) = \sum_{A=1}^N \frac{Z_A}{|r - R_A|} - \int \frac{\rho(r')d^3r'}{|r - r'|} \quad (1)$$

where N denotes the total number of nuclei in the molecule, and the two terms refer to the bare nuclear potential and the electronic contributions, respectively. The balance of these two terms gives the effective localization of electron-rich regions in the molecular system. Molecular frontier orbitals HOMO and LUMO as well as MESP of all optimized structures were visualized with Maestro (version 9.1, Schrodinger package [28]). The outlines of these MESP maps provide a measure of the overall size of the molecule and the color-coded surface gives the locations of negative and positive electrostatic potentials.

Results and discussion

Effects of noscapinoids on proliferation of cancer cells

The absolute anti-cancer activity (IC₅₀ value) of noscapinoids against human lymphoblastoid cells (CEM) were measured by MTS assay. The negative log of IC₅₀ value (expressed as pIC₅₀) is used as anti-cancer activity in evaluating the structure–activity relationships of noscapinoids quantitatively. Given a normal distribution (bell-shaped curve) of activity values of a set of compounds, the rule of thumb in developing fairly accurate QSAR models suggests that the difference between the highest and lowest biological activity of the compounds should be three to four

Table 2 Experimental and predicted activities for training and test set compounds based on 3D quantitative structure–activity relationship (QSAR) models along with their calculated electronic features. pIC_{50} $-(\log IC_{50})$, E_{solv} solvation energy, $LUMO$ lowest unoccupied molecular orbital, $HOMO$ highest occupied molecular orbital, HLG HOMO–LUMO energy gap

Sl. no.	Experimental pIC_{50} (M)	Predicted pIC_{50} (M)	Residual	E_{solv} (kcal/mol)	HOMO (eV)	LUMO (eV)	HLG (eV)	QM dipole (Debye)
1	4.780	4.970	0.190	-51.82	-0.214	-0.037	-0.177	13.61
2	4.548	4.770	0.222	-43.60	-0.211	-0.015	-0.196	12.49
3 ^a	5.921	5.400	0.521	-79.32	-0.229	-0.078	-0.151	13.38
4	4.345	4.455	0.110	-47.54	-0.212	-0.023	-0.189	12.20
5	5.721	5.664	0.057	-83.65	-0.225	-0.076	-0.149	13.45
6	5.553	5.320	0.233	-43.34	-0.211	-0.066	-0.176	12.15
7	5.638	5.440	0.198	-52.70	-0.219	-0.069	-0.150	13.77
8 ^a	4.810	4.880	0.070	-47.91	-0.213	-0.019	-0.194	12.13
9 ^a	5.585	5.680	0.095	-55.02	-0.207	-0.062	-0.145	13.07
10	5.585	5.770	0.185	-47.37	-0.204	-0.065	-0.139	11.88
11 ^a	5.520	5.300	0.220	-54.40	-0.214	-0.057	-0.157	14.89
12	5.523	5.413	0.110	-49.69	-0.218	-0.059	-0.159	13.61
13	5.000	4.937	0.063	-57.14	-0.217	-0.088	-0.129	13.08
14	5.000	5.160	0.160	-51.48	-0.212	-0.086	-0.126	11.71
15	4.410	4.600	0.190	-50.96	-0.221	-0.044	-0.177	13.03
16	4.516	4.710	0.194	-43.84	-0.215	-0.042	-0.173	12.29
17	4.319	4.170	0.149	-14.65	-0.207	-0.045	-0.162	6.14
18 ^a	4.345	4.230	0.115	-14.77	-0.207	-0.036	-0.171	6.55
19	4.355	4.184	0.171	-13.73	-0.207	-0.024	-0.183	5.99
20	4.392	4.211	0.181	-10.62	-0.207	-0.044	-0.163	5.86
21	4.378	4.260	0.118	-19.15	-0.209	-0.045	-0.164	5.98
22 ^a	4.346	4.300	0.046	-18.31	-0.209	-0.025	-0.184	6.40
23 ^a	4.333	4.300	0.033	-17.19	-0.208	-0.042	-0.166	5.67
24	4.353	4.240	0.113	-19.41	-0.209	-0.043	-0.166	5.81
25 ^a	4.356	4.310	0.046	-16.20	-0.206	-0.048	-0.158	5.80
26	4.286	4.166	0.120	-27.55	-0.214	-0.035	-0.179	14.60
27	4.403	4.350	0.053	-39.51	-0.214	-0.046	-0.168	14.18
28 ^a	4.371	4.550	0.179	-31.40	-0.216	-0.031	-0.185	14.01
29	4.387	4.250	0.137	-38.09	-0.215	-0.049	-0.166	14.23
30	4.374	4.221	0.153	-34.14	-0.221	-0.050	-0.171	14.48
31	4.450	4.211	0.239	-34.64	-0.220	-0.039	-0.181	14.24
32	4.271	4.330	0.059	-39.63	-0.211	-0.016	-0.195	10.59
33 ^a	4.252	4.410	0.158	-37.78	-0.205	-0.011	-0.194	5.48
34	4.277	4.400	0.123	-37.32	-0.211	-0.017	-0.194	10.49
35	4.427	4.348	0.079	-24.37	-0.214	-0.035	-0.179	12.81
36	4.364	4.184	0.180	-11.72	-0.207	-0.036	-0.171	3.59

^a Compounds in the test set

orders of magnitude [36]. However, with the array of noscapioids examined in this study, the distribution of activities values was skewed to the left as the majority of compounds (22 out of 36) have weaker activity ($pIC_{50} < 4.5$ M). Given this asymmetric distribution of activity, the actual range of raw experimental IC_{50} data in our case is, in fact, quite impressive (ranging from 1.2 to 56 μ M, Table 1)

and can be reasonably used to assess the structure–activity relationship among noscapioids. Noscapiine derivatives **3**, **5–7**, and **9–14** have significantly better activities ($pIC_{50} \geq 5.0$ M) than the other compounds. On the contrary, derivatives **17–36** (aryl substituted N-carbamoyl/N-thiocarbamoyl noscapiine analogues) generally showed very weak or no activity.

Table 3 Summary of QSAR results for five best common pharmacophore hypotheses (CPHs) with survival scores and statistical parameters. *SD* Standard deviation of the regression, R^2 value of R^2 for the regression, F -value variance ratio, P probability value of significance, *RMSE* residual mean square error between experimental

	CPH1 (AAHHPRR)	CPH2 (AAHHPRR)	CPH3 (AHHHPRR)	CPH4 (AAHHPRR)	CPH5 (AHHHPRR)
Survival score	13.489	13.466	13.464	13.412	13.405
Survival inactive	11.556	11.456	11.526	11.468	11.501
SD	0.250	0.268	0.260	0.257	0.254
R^2	0.912	0.882	0.795	0.799	0.804
F -value	31.400	26.300	28.400	29.200	30.000
P	3.97e-08	1.82e-07	9.58e-08	7.38e-08	5.86e-08
RMSE	0.215	0.258	0.251	0.322	0.262
Q^2	0.908	0.835	0.844	0.743	0.830
Pearson- R	0.951	0.938	0.953	0.899	0.931

Pharmacophore perception

The 36 noscapioids considered as potential inhibitors of cell proliferation were divided randomly into an active set ($pIC_{50} > 5.0$) comprised of 10 compounds (3, 5–7, 9–14) and an inactive set ($pIC_{50} < 4.5$) comprised of 23 compounds (4, 15, 17–36); the remaining 3 compounds (2, 6, 8) were not classified in these categories and are termed the “intermediate” class. To determine the molecular components required for anti-cancer activity of noscapioids, we generated various pharmacophore hypotheses containing five, six and seven sites using a terminal box size of 1.0 Å. Based on the molecular occupancy of the pharmacophore features, CPHs composed of five and six sites were rejected. A total of 1,838 CPHs consisting of seven sites belonging to 14 types (AAHHHPR, AAAAHP, AAAHHHR, AAAAHRR, AAAAHHP, AAAHHHP, AAHHPRR, AAAHHPR, AAAHPRR, AAAHHRR, AAAAHR, AAAAPRR, AHHHPRR and AAHHHRR) were subjected to stringent scoring function analysis with respect to active molecules and then with inactive molecules. Default parameters for site, vector, and volume were used for scoring CPHs. In addition, ligand activity (expressed as pIC_{50}) was incorporated with a weight of 1.0. The hypotheses that survived the scoring process were examined further by developing atom-based 3D-QSAR models. Table 3 summarizes the statistical data of the top ranked CPHs (with their survival scores) labeled as CPH1–CPH5. Among these, CPH1 was found to be the best based on different statistical parameters such as R^2 , Q^2 , Pearson- R , SD, RMSE and F -value. For example, R^2 and Q^2 values of 0.912 and 0.908 corroborate with the criteria for a QSAR model to be highly predictive. This yielded a good Pearson- R value of 0.951 and high F -value of 31.40. The root mean square error for the model was 0.215, which is an indicator of the quality of fit between the experimental and predicted

and predicted activities, Q^2 cross validated R^2 using leave one out cross validation technique for the predicted activities, Pearson- R , correlation between the predicted and experimental activity for the test set

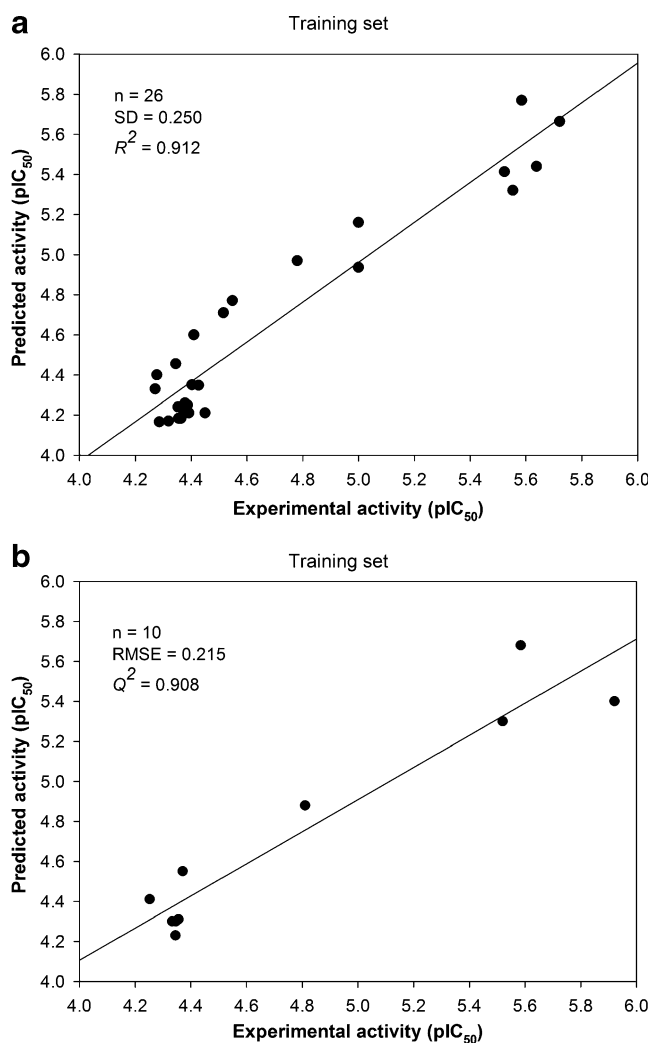


Fig. 2 Relationship between experimental and predicted activities (pIC_{50}) of **a** training set and **b** test set compounds based on 3D quantitative structure–activity relationship (QSAR) model; $R^2=0.920$; $Q^2=0.885$ and Pearson- $R=0.951$

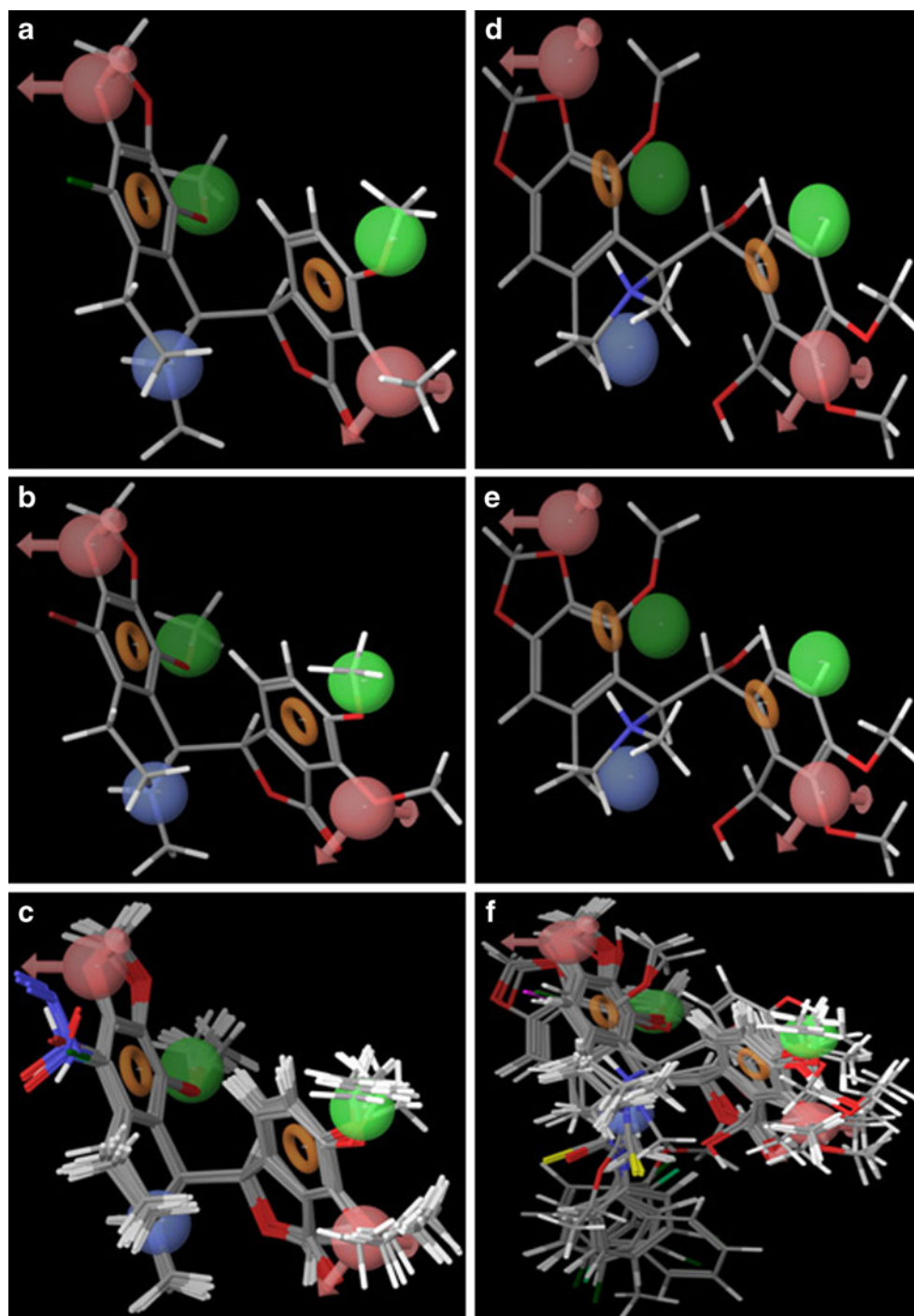
pIC₅₀ values (Fig. 2). The predicted and the experimental values of the training and test set molecules are given in Table 2. The most successful hypothesis—CPH1—was retained for further studies consisting of the following molecular components (Fig. 3):

1. Two acceptor features (A) mapping on the lone pair of vectors of oxygen atom (methoxy group) attached to dimethoxy isobenzofuranone ring (lower section of noscapine in Fig. 1).

2. Two aromatic ring features (R) mapping on the benzene rings.
3. Two hydrophobic group features (H) corresponding to two methyl groups
4. One positive group feature (P) corresponding to the nitrogen atom of the isoquinoline ring.

All highly active compounds of the noscapinoid family map all the functional features of the best predictive hypothesis (CPH1). For example, the two most potent

Fig. 3 Top ranked pharmacophore model (CPH1) mapped onto **a** the fittest compound **3**; **b** the most active compound **5**; **c** all the active compounds; on the inactive compounds **d 32**; **e 33** as well as **f** with all the inactive compounds. Pharmacophore variants considered for anti-cancer activity were composed of seven sites with the following chemical features: hydrogen-bond acceptor (A, pink sphere), aromatic ring (R, ring), hydrophobic (H, green sphere), and positively charged group (P, blue sphere)



members of the series—compounds **3** (9-chloro-noscapine) and **5** (9-bromo-noscapine)—map perfectly onto all the pharmacophoric features of hypothesis CPH1 (Fig. 3a,b). The other eight active compounds also map well onto the pharmacophore hypothesis (Fig. 3c). In contrast, the most inactive compounds **32** and **33** do not map onto the pharmacophore hypothesis (Fig. 3d,e), whereas some others map only partially (to only a few of the features of CPH1). The acceptor and polar groups seem to be crucial for anti-cancer activity, because these groups do not map in less active analogues. Also, in some of the less active analogues, the aromatic rings do not map properly, in spite of good mapping of the acceptor and positive group features (Fig. 3f). This stresses the equal importance of the aromatic rings in anti-cancer activity. Thus, overall, our complete pharmacophore model can distinguish accurately the active and inactive compounds of the noscapinoid family.

Lowest unoccupied and highest occupied molecular orbitals

Frontier orbital electron densities on molecules provide a useful means of characterizing donor–acceptor interactions in detail. It has been shown that these orbitals play a major role in governing many chemical reactions. For example, they are responsible for the formation of many charge transfer complexes. The energy of the HOMO is related directly to the ionization potential and it characterizes the susceptibility of the molecule towards an attack by electrophiles. On the other hand, the LUMO is related with electron affinity and gives an idea of the susceptibility of the molecule towards attack by nucleophiles. Among the noscapinoids used in this study, HOMO and LUMO energies are both small, ranging between -0.205 to -0.227 eV and -0.011 to -0.088 eV, respectively, indicating the fragile nature of the bound electrons (Table 2). The small energy gap between HOMO and LUMO energies (varies between 0.05 and 0.196 eV) makes both rapid electron transfer and exchange equally possible, making these compounds very reactive. Noscapinoids with higher anti-cancer activity ($\text{pIC}_{50} > 5.0$) are those with more negative LUMOs values (on average, -0.071 eV) in comparison to the less active compounds (on average, -0.035 eV). Most active compounds can be distinguished clearly from less active compounds on the basis of their LUMO energies, as evident from the scatter plot of LUMO energies against activity (Fig. 4). Thus, manipulating the functional groups that affect LUMO energies could have a direct impact on the anti-cancer activity of noscapinoids. LUMO sites plotted onto the active molecules **1**, **3**, **5**, **7**, **11**, **12** were scattered over the dimethoxy and the isobenzofuranone (lower section) ring; among compounds **9** and **10**, LUMO sites were located at the azido groups; whereas for compounds **13** and **14**, they are distributed over the nitro

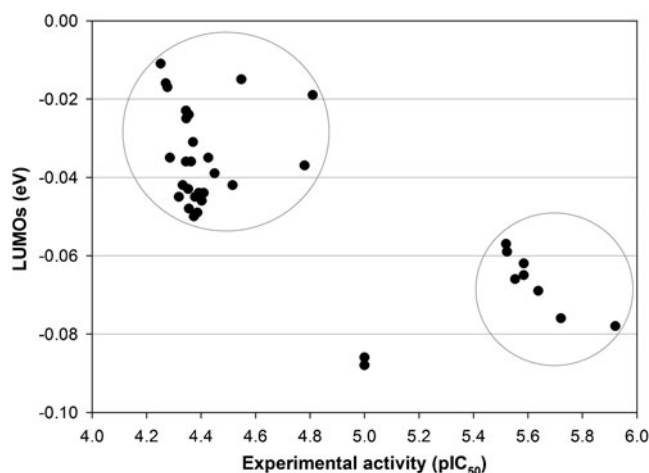


Fig. 4 Relationship between experimental activity (pIC_{50}) versus lowest unoccupied molecular orbital (LUMO) energy (eV) of the compounds. The active compounds ($\text{pIC}_{50} > 5.0$) are clearly separated from the inactive compounds ($\text{pIC}_{50} < 4.5$)

group, imply the susceptibility of these groups toward nucleophilic attack (Fig. 5). In contrast, among the least active noscapinoid compounds (**2**, **4**, **15**, **16**, **17–36**), such characteristic LUMO sites were missing. In striking contrast, HOMOs are located mainly at the isoquinoline ring in this series of noscapinoids.

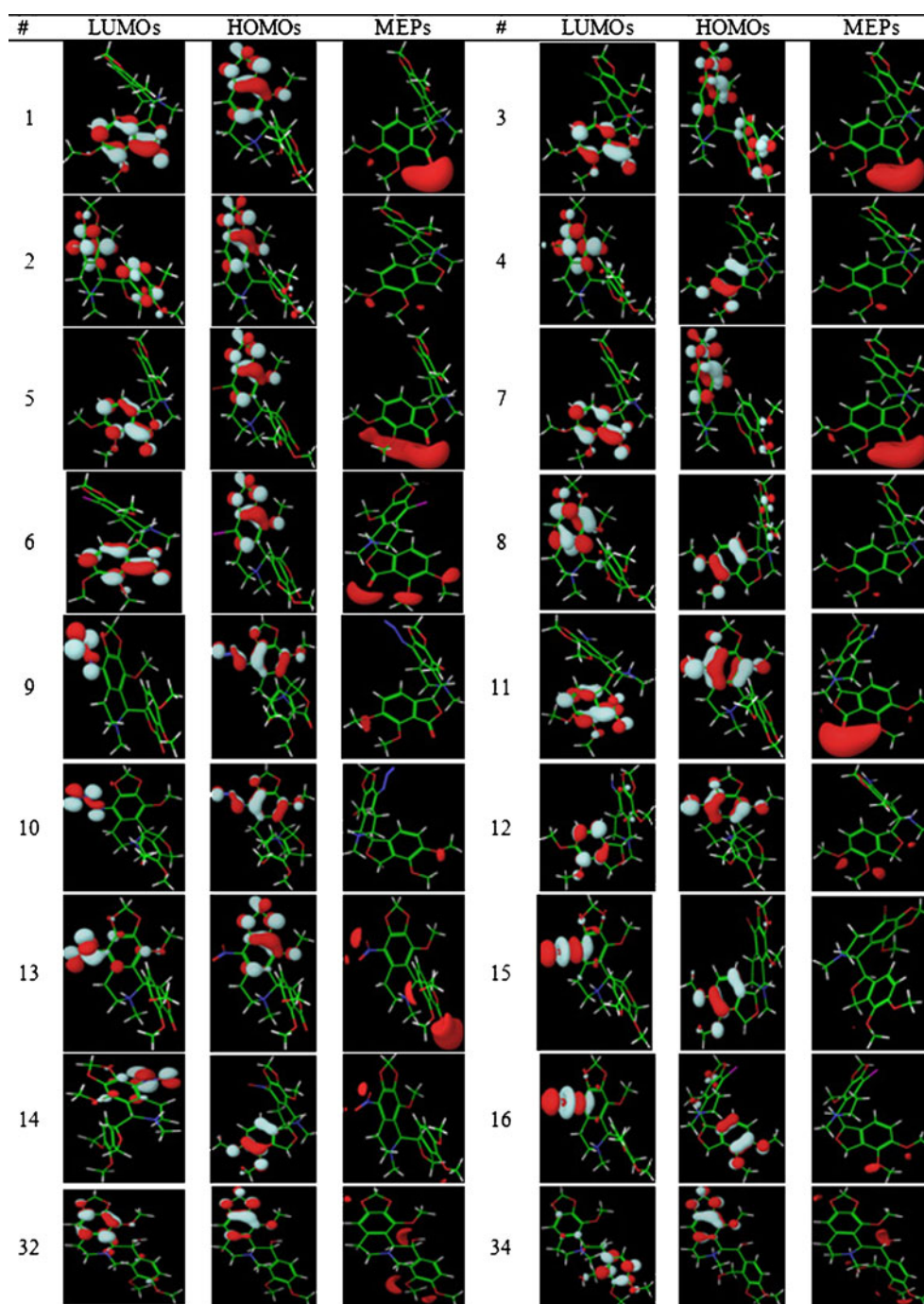
Electrostatic potential profiles

Our results show that all the most active noscapinoids share specific electronic properties and are different from the least active compounds (Fig. 5). In our classification, all “most active compounds” showed an electronegative potential region (red color) near the oxygen atoms of the furanone ring, extending laterally over the oxygen atom of the dimethoxy group (L-shaped). Among the nitro derivatives (**13** and **14**), one more prominently localized negative charged region was seen near the oxygen atom of the nitro group. In the least active compounds, most negative potentials due to the oxygen atoms are missing. It is apparent from these results that the negative potentials near the oxygen atom of the dimethoxy isobenzofuranone ring and nitro group are crucial for activity. These electrostatic potential (ESP) profiles generated above are consistent with the pharmacophore model in which the hydrogen bond acceptor site is located at the oxygen atom of the methoxy isobenzofuranone ring.

Dipole moment

Dipole moments of the noscapinoids varied from 5.48 to 14.89 Debye (Table 2). All the most active compounds had higher dipole moment (> 11 Debye). However, some of the

Fig. 5 Three dimensional LUMOs, highest occupied molecular orbitals (HOMOs) and electrostatic potential profiles of the optimized structures for most active (**3**, **5–7**, **9–14**), intermediate (**1**, **2**, **8**, **16**) and inactive (**4**, **15**, **32**, **34**) compounds. The orientations of the molecular structures are adjusted for better visualization of calculated molecular orbitals and electrostatic potential profiles



less active compounds also had high dipole moment values (for example, compounds **26–31** have a dipole moment of ~14 Debye) and hence there is no clear correlation between dipole moment and biological activity.

Aqueous stabilization

Aqueous solubility is one of the most important properties contributing to the oral bioavailability of a

drug. A value of $50 \mu\text{g ml}^{-1}$ is considered as the acceptable cut off for most preclinical drug development programs [37]. Solvation energy gives a measure of compound solubility, in that higher negative values of solvation free energy are indicative of higher aqueous solubility. In our noscapine derivative family, the solvation free energies of noscapinoids varied from $-10.62 \text{ kcal mol}^{-1}$ (**20**) to $-83.63 \text{ kcal mol}^{-1}$ (**5**) (Table 2) indicating the good aqueous solubility of these com-

pounds. The most active compounds have solvation energy $> -50 \text{ kcal mol}^{-1}$, whereas inactive molecules—mostly aryl derivatives (17–36)—have less negative solvation energy ($< -40 \text{ kcal mol}^{-1}$). Although, we did not pursue quantitative studies on the water solubility of noscapinoids, we routinely prepared stock solutions in ethanol or dimethylsulfoxide that were then diluted 100 to 1,000 fold in aqueous solution to determine the biological activity of noscapinoids. Furthermore, in our previous studies, these compounds were administered by feeding noscapine orally in aqueous solutions, or even by simply dissolving in drinking water [38]. Satisfactorily, these compounds also demonstrated favorable pharmacokinetics (clearance in 6–10 h) [39], consistent with the general belief that water-soluble drugs also display good pharmacokinetic behavior. Hence, the solvation energies computed in this study can be used to guide pharmacokinetic optimization of the compounds under study.

Conclusions

This body of work represents the first report of a highly predictive pharmacophore model of noscapine and its derivatives (the noscapinoids). A total of 1,838 different hypotheses were developed and examined. A final seven-point pharmacophore with the common features of a two hydrogen bond acceptor, two aromatics rings, two hydrophobic groups, and a positively charged group was associated with the most significant QSAR model. Calculated quantum chemical properties provide additional mechanistic details about the activity of these compounds. This work revealed that the presence of most negative potential regions above the oxygen atoms of the dimethoxy isobenzofuranone ring and the nitro group are essential for potent anticancer activity. The most active compounds are associated with higher negative values of LUMO energies. Aqueous solvation energies values give an indication of the aqueous solubility and can be used as a guide to pharmacokinetic optimization of these molecules. Modification of electronic properties in accordance with the results of this work by incorporating such structural features into the scaffold structure of noscapine will certainly achieve the final goal of improved anti-cancer noscapinoids.

Acknowledgments We thank Dr. William Beck for providing the drug-resistant CEM cell line used in this study and for advice. We are indebted to the anonymous reviewers of this manuscript for helpful suggestions. Grant support: National Institutes of Health (National Institute of Cancer) grants CA-095317-01A2 (H.C.J.) and Better Opportunities for Young Scientists in Chosen Areas of Science and Technology fellowship (SR/BY/L-37/09; Department of Science and Technology, Government of India) to P.K.N.

References

- Chopra RN, Mukherjee BI, Dikshit BB (1930) Narcotine: its pharmacological action and therapeutic uses. *Indian J Med Res* 18:35–49
- Winter CA, Flataker L (1954) Antitussive compounds testing methods and results. *J Pharmacol Exp Ther* 112:99–108
- Idanpaan-Heikkila JE, Jalonen K, Vartiainen A (1967) Evaluation of the antitussive effect of noscapine and codeine on citric acid cough in guinea-pigs. *Acta Pharmacol Toxicol* 25:333–338
- Ye K, Ke Y, Keshava N, Shanks J, Kapp JA, Tekmal RR, Petros J, Joshi HC (1998) Opium alkaloid noscapine is an antitumor agent that arrests metaphase and induce apoptosis in dividing cells. *Proc Natl Acad Sci USA* 95:1601–1606
- Ye K, Zhou J, Landen JW, Bradbury EM, Joshi HC (2001) Sustained activation of p34(cdc2) is required for noscapine-induced apoptosis. *J Biol Chem* 276:46697–46700
- Zhou J, Gupta K, Yao J, Ye K, Panda D, Giannakakou P, Joshi HC (2002) Paclitaxel-resistant human ovarian cancer cells undergo c-Jun N-terminal kinase-mediated apoptosis in response to noscapine. *J Biol Chem* 277:39777–39785
- Zhou J, Panda D, Landen JW, Wilson L, Joshi HC (2002) Minor alteration of microtubule dynamics causes loss of tension across kinetochore pairs and activates the spindle checkpoint. *J Biol Chem* 277:17200–17208
- Zhou J, Gupta K, Aggarwal S, Aneja R, Chandra R, Panda D, Joshi HC (2003) Brominated derivatives of noscapine are potent microtubule-interfering agents that perturb mitosis and inhibit cell proliferation. *Mol Pharmacol* 63:799–807
- Aneja R, Vangapandu SN, Lopus M, Chandra R, Panda D, Joshi HC (2006) Development of a novel nitro-derivative of noscapine for the potential treatment of drug-resistant ovarian cancer and T-cell lymphoma. *Mol Pharmacol* 69:1801–1809
- Aneja R, Vangapandu SN, Lopus M, Visweswarappa VG, Dhiman N, Verma A, Chandra R, Panda D, Joshi HC (2006) Synthesis of microtubule-interfering halogenated noscapine analogs that perturb mitosis in cancer cells followed by cell death. *Biochem Pharmacol* 72(4):415–426
- Aneja R, Vangapandu SN, Joshi HC (2006) Synthesis and biological evaluation of a cyclic ether fluorinated noscapine analog. *Bioorg Med Chem* 14:8352–8358
- Zhou J, Liu M, Luthra R, Jones J, Aneja R, Chandra R, Tekmal RR, Joshi HC (2005) EM012, a microtubule-interfering agent, inhibits the progression of multidrug-resistant human ovarian cancer both in cultured cells and in athymic nude mice. *Cancer Chemother Pharmacol* 55:461–465
- Aggarwal S, Ghosh NN, Aneja R, Joshi HC, Chandra R (2002) A convenient synthesis of aryl-substituted N-carbamoyl/N-thiocarbamoyl narcotine and related compounds. *Helv Chim Acta* 85:2458–2462
- Dixon SL, Smondyrev AM, Rao SN (2006) PHASE: A novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des* 67:370–372
- Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner R (2006) PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening. 1. Methodology and preliminary results. *J Comput Aided Mol Des* 20:647–671
- Narkhede SS, Degani MS (2007) Pharmacophore refinement and 3D-QSAR studies of histamine H3 antagonists. *QSAR Comb Sci* 26:744–753
- Tawari NR, Bag S, Degani MS (2008) Pharmacophore mapping of a series of pyrrolopyrimidines, indolopyrimidines and their congeners as multidrug resistance-associated protein (MRP1) modulators. *J Mol Model* 14:911–921

18. Bag S, Tawari NR, Degani MS (2009) Insight into inhibitory activity of *Mycobacterial* dihydrofolate reductase inhibitors by in silico molecular modeling approaches. *QSAR Comb Sci* 28:296–311
19. Politzer P, Murray JS, Peralta-Inga Z (2001) Molecular surface electrostatic potentials in relation to noncovalent interactions in biological systems. *Int J Quantum Chem* 85:676–684
20. Politzer P, Murray JS, Concha MC (2002) The complementary roles of molecular surface electrostatic potentials and average local ionization energies with respect to electrophilic processes. *Int J Quantum Chem* 88:19–27
21. Gejji SP, Suresh CH, Babu K, Gadre SR (1999) Ab initio structure and vibrational frequencies of $(CF_3SO_2)_2N^+Li^-$ ion pairs. *J Phys Chem* 103:7474–7480
22. Bhattacharjee AK, Karle JM (1998) Functional correlation of molecular electronic properties with potency of synthetic carbinolamines antimalarial agents. *Bioorg Med Chem* 6:1927–1933
23. Bhattacharjee AK, Karle JM (1999) Stereoelectronic properties of antimalarial artemisinin analogues in relation to neurotoxicity. *Chem Res Toxicol* 12:422–428
24. Vijayalakshmi KP, Suresh CH (2008) Role of structural water molecule in HIV protease-inhibitor complexes: a QM/MM study. *J Comput Chem* 29:1840–1849
25. Tomasi J, Bonaccorsi R, Cammi R (1990) In: Maksic R (ed) *Theoretical models of chemical bonding*. Springer, Berlin, pp 230–268
26. Beck WT, Cirtain MC (1982) Continued expression of vinca alkaloid resistance by CCRF-CEM cells after treatment with tunicamycin or pronase. *Cancer Res* 42:184–189
27. Skehan P, Storeng R, Scudiero D, Monks A, McMahon J, Vistica D, Warren JT, Bokesch H, Kenney S, Boyd MR (1990) New colorimetric anti-cancer activity assay for anti-cancer-drug screening. *J Natl Cancer Inst* 82:1107–1112
28. Schrödinger, LLC Phase, Version 3.2. Schrödinger, LLC, New York, NY; <http://www.schrodinger.com>
29. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 37:785–789
30. Becke AD (1993) A new mixing of Hartree-Fock and local density-functional theories. *J Chem Phys* 98:1372–1377
31. Binkley JS, Pople JA, Hehre WJ (1980) Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *J Am Chem Soc* 102:939–947
32. Gordon MS, Binkley JS, Pople JA, Pietro WJ, Hehre WJ (1982) Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements. *J Am Chem Soc* 104:2797–2803
33. Pietro WJ, Francl MM, Hehre WJ, Defrees DJ, Pople JA, Binkley JS (1982) Self-consistent molecular orbital methods. 24. Supplemented small split-valence basis sets for second-row elements. *J Am Chem Soc* 104:5039–5048
34. Rodrigues T, dos Santos DJVA, Moreira R, Guedes RC (2010) A quantum mechanical study of novel potential inhibitors of cytochrome bc₁ as antimalarial compounds. *Int J Quantum Chem* 111:1196–1207. doi:10.1002/qua.22741
35. Tawari NR, Degani MS (2009) Pharmacophore mapping and electronic feature analysis for a series of nitroaromatic compounds with antitubercular activity. *J Comput Chem* 31:739–751
36. Checchi PM, Nettles JH, Zhou J, Snyder JP, Joshi HC (2003) Microtubule-interacting drugs for cancer treatment. *Trends Pharmacol Sci* 24:361–365
37. Budha NR, Mehrotra N, Tangallapally R, Rakesh QJ, Daniels AJ, Lee RE, Meibohm B (2008) Pharmacokinetically-guided lead optimization of nitrofuranyl amide anti-tuberculosis agents. *AAPS J* 10:157–165
38. Landen JW, Lang R, McMahon SJ, Rusan NM, Yvon AM, Adams AW, Sorcinelli MD, Campbell R, Bonaccorsi P, Ansel JC, Joshi HC (2002) Noscapine alters microtubule dynamics in living cells and inhibits the progression of melanoma. *Cancer Res* 62:4109–4114
39. Dahlstrom B, Mellstrand T, Lofdahl CD, Hohansson M (1982) Pharmacokinetic properties of noscapine. *Eur J Clin Pharmacol* 22:535–539

A density functional theory approach toward substituent effect in Meerwein–Eschenmoser–Claisen rearrangement

Rahim Ghadari · Ahmad Shaabani

Received: 3 August 2010 / Accepted: 29 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract B3LYP/6-31 G(d) level of theory has been used for the examination of substituent effect in the concerted step of the Meerwein–Eschenmoser–Claisen rearrangement. In this regard, the effect of NO₂ and NH₂ groups in different positions has been investigated. The obtained results show that substituent effect is very sensitive to its position and configuration. Electron withdrawing and electron donating groups in different positions and various configurations show different and sometimes opposite results.

Keywords B3LYP/6-31 G(d) · Concerted procedure · Electron donating groups · Electron withdrawing groups · Meerwein–Eschenmoser–Claisen rearrangement

Introduction

Meerwein–Eschenmoser–Claisen rearrangement (MEC-R) which is well-known as Eschenmoser–Claisen rearrangement is one of the most useful reactions of organic chemistry. This reaction is one of the well-known and useful subdivisions of the Claisen rearrangement [1]. This reaction involves conversion of an allylic alcohol **1** to a ketene *N,O*-acetal **3**, which undergoes a rapid [3,3]-sigmatropic rearrangement to yield a γ - δ -unsaturated amide **4** (Scheme 1). Although this reaction was introduced by Meerwein in 1961 [2], until

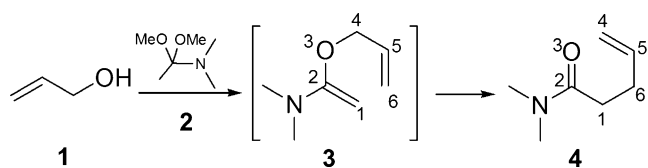
1964 when Eschenmoser [3] started his experimental investigations on it, this reaction had not been widely used in organic synthesis.

Because of the presence of an electron donating amino group in the MEC-R, the rate of the MEC-R in the pericyclic step is more than Claisen rearrangement. In comparison with Claisen or Johnson–Claisen rearrangement, MEC-R has higher yield and better selectivity [4]. Because the reaction conditions are neutral, this rearrangement can be used for sensitive substrates. Due to the highly ordered 6-membered transition state, this rearrangement could proceed with good stereo control.

Because of the above mentioned advantages, MEC-R had been used in different synthetic procedures. Wovkulich and coworkers used MEC-R in the total synthesis of Mevinolin [5]. Total synthesis of enokipodins A and B has been carried out by Yoshida and coworkers using MEC-R [6]. This rearrangement is one of the key steps in the synthesis of pyrethroid amides *via* epoxy amide cyclization which has been reported by Majewski and Snieckus [7]. It has been reported that Claisen rearrangement could be used for the synthesis of C-Glycopyranosides, but the overall process is unsatisfactory because of the low conversion to the vinyl ether intermediate. The best results could be obtained using MEC-R [8]. The synthesis of anti- β -substituted γ , α -unsaturated amino acids *via* MEC-R have been reported too [9]. Although Kazmaier–Claisen rearrangement could be used for this reaction, Kazmaier's strategy does not work well for racemic anti- β -substituted γ , α -unsaturated amino acids. Optically active trichothecane skeleton has been synthesized using MEC-R [10]. Pepluanin A and euphosalicin have been synthesized using MEC-R [11]. This rearrangement has been used widely in the total synthesis of natural products [12–15].

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1080-x) contains supplementary material, which is available to authorized users.

R. Ghadari · A. Shaabani (✉)
Department of Chemistry, Shahid Beheshti University, G. C.,
P. O. Box 19396–4716, Tehran, Iran
e-mail: a-shaabani@cc.sbu.ac.ir



Scheme 1 Meerwein–Eschenmoser–Claisen rearrangement

Catalytic MEC-R has been investigated by various research groups. Use of palladium complexes as an asymmetric catalyst in the MEC-R has been reported by Linton and Kozłowski [16]. Between different ligands which have been used in this investigation, the bisphosphines and phosphinoxazolines have excellent yields and higher enantioselectivity. Tetra-*n*-butylammonium fluoride has been used as a catalyst in MEC-R, too [17].

Although a computational study on the Claisen rearrangement has been reported by different research groups [18–29], to the best of our knowledge, a full computational study has not yet been carried out on the MEC-R. Therefore, in this work, we carried out a systematic study regarding the effect of electron donating and electron withdrawing substituents on this rearrangement. The effect of these groups on the transition state (TS) and feasibility of

the reaction have been investigated. This rearrangement has different steps which in the present work, the highly ordered 6-membered transition state has been investigated. In this study the chair conformation has been selected as the preferable structure for TS [28, 29]. Our main aim is to bring an insight into the effect of substituent on the structure of TSs and energy requirements of MEC-R.

Computational methods

B3LYP/6-31 G(d) computational level of theory was used for geometry optimizations. In the case of reactants and products, the minima have been confirmed by the absence of imaginary frequency. The nature of the transition structures was confirmed by the presence of a single imaginary frequency (NIMG=1). All the transition state structures were traced with IRC calculation in both directions. The end points of IRC calculations were optimized as the structure of reactant and product. The non-scaled frequencies were used to compute zero-point energies. All of the calculations were carried out in gas phase employing Gaussian 98 [30]. GaussView 3.0 [31] and CYLview (v1.0beta) [32] programs were used as graphical interfaces.

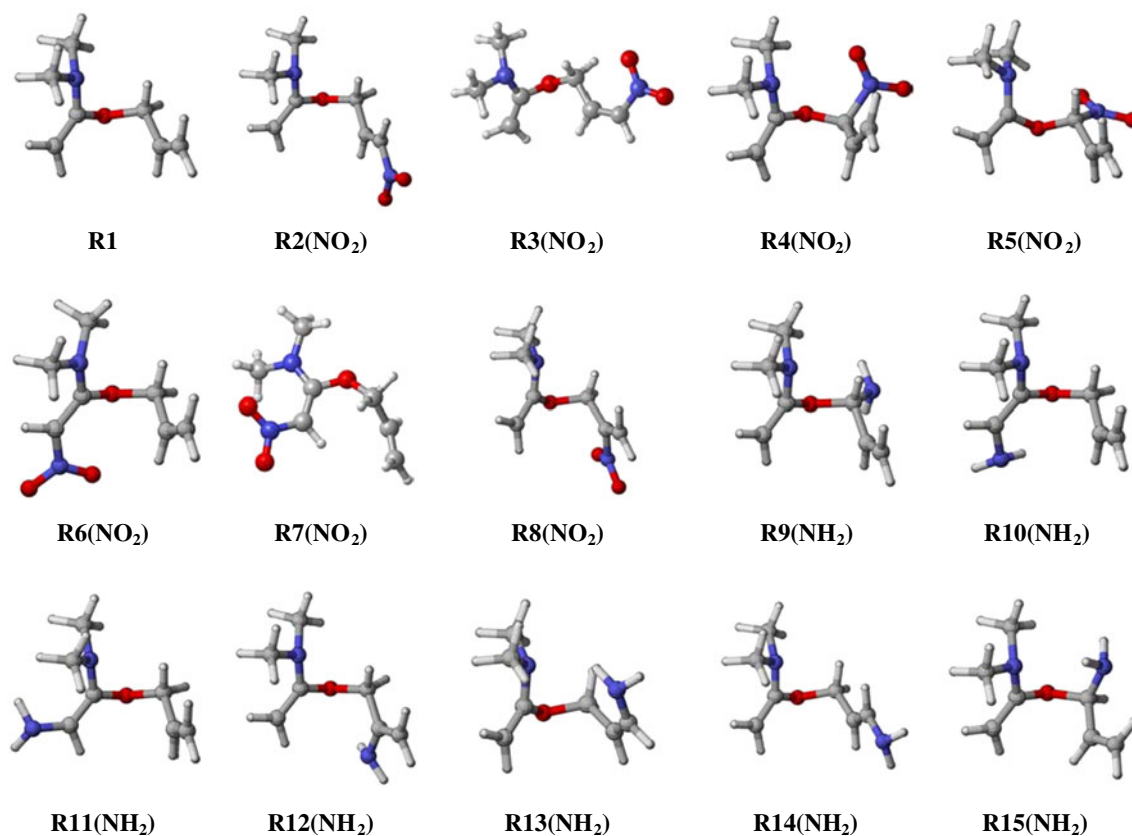


Fig. 1 Investigated optimized minima structures as reactant

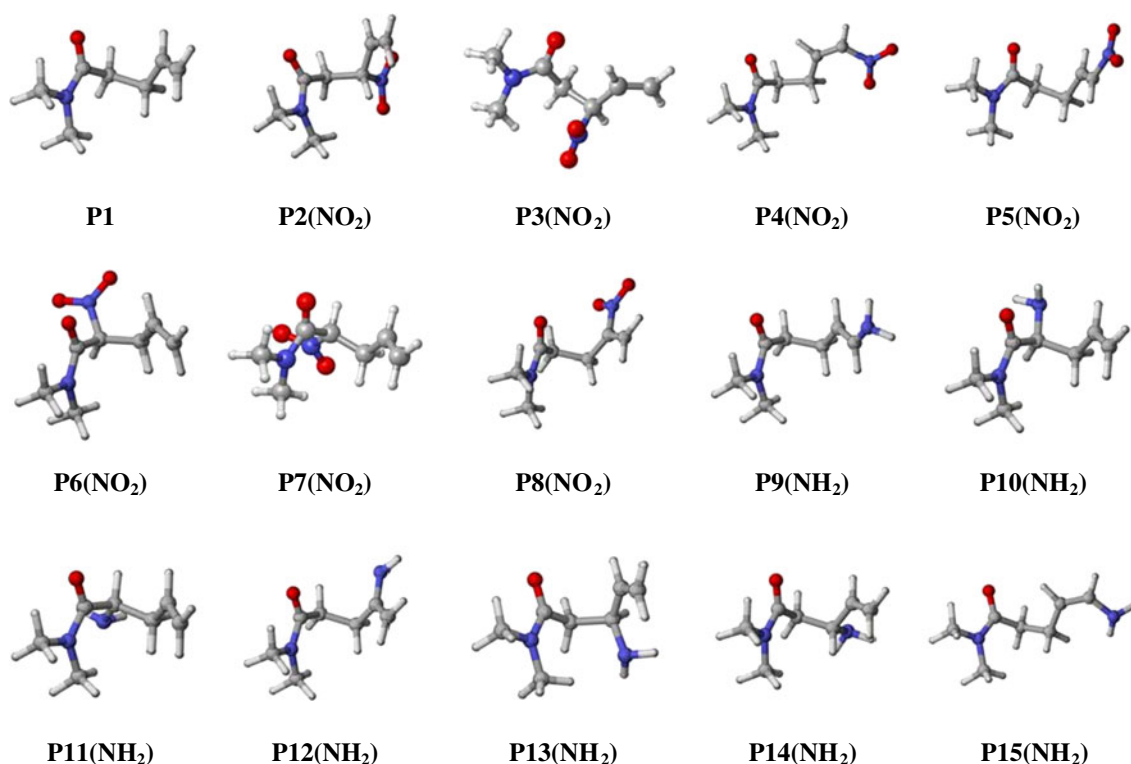


Fig. 2 Investigated optimized minima structures as product

Results and discussion

The effect of NH_2 and NO_2 substitutions in different positions has been investigated (Scheme 1). All of the reactants, transition states (TSs), and products were fully optimized. In the case of reactants (Fig. 1) and products (Fig. 2), the end point of IRC calculations have been selected for geometry optimization. The chair conformation

was assumed as the most favorable conformation in all of the transition structures.

Structures of reactants

The optimized structures of reactants were presented in Fig. 1. The dihedral angles ($\text{C}2\text{-O-C}4\text{-C}5$) of the optimized structures of reactants and selected bond lengths have been

Table 1 Selected bond lengths and dihedral angles of reactants^a

Structure	R	C2-O (Å)	O-C4 (Å)	C4-C5 (Å)	C2-O-C4-C5 (°)
R1	–	1.38	1.44	1.50	75.2
R2(NO_2)	6- NO_2 (<i>E</i>)	1.39	1.43	1.50	75.2
R3(NO_2)	6- NO_2 (<i>Z</i>)	1.38	1.43	1.50	76.8
R4(NO_2)	4- NO_2 (<i>R</i>)	1.40	1.40	1.51	13.3
R5(NO_2)	4- NO_2 (<i>S</i>)	1.39	1.40	1.49	97.8
R6(NO_2)	1- NO_2 (<i>Z</i>)	1.35	1.46	1.50	58.1
R7(NO_2)	1- NO_2 (<i>E</i>)	1.36	1.45	1.50	94.5
R8(NO_2)	5- NO_2	1.38	1.44	1.50	80.2
R9(NH_2)	4- NH_2 (<i>S</i>)	1.38	1.47	1.50	78.0
R10(NH_2)	1- NH_2 (<i>Z</i>)	1.39	1.44	1.50	66.7
R11(NH_2)	1- NH_2 (<i>E</i>)	1.39	1.44	1.50	73.6
R12(NH_2)	5- NH_2	1.38	1.43	1.51	81.3
R13(NH_2)	6- NH_2 (<i>Z</i>)	1.37	1.46	1.49	66.4
R14(NH_2)	6- NH_2 (<i>E</i>)	1.37	1.46	1.49	68.8
R15(NH_2)	4- NH_2 (<i>R</i>)	1.38	1.48	1.51	53.5

^a Absolute configuration of structures have been shown in the parentheses. For atom numbering see Scheme 1

Table 2 Selected bond lengths and dihedral angles of products^a

Structure	R	C1-C2 (Å)	C1-C6 (Å)	C5-C6 (Å)	C2-C1-C6- C5 (°)
P1	–	1.53	1.55	1.50	–55.9
P2(NO ₂)	6-NO ₂ (<i>R</i>)	1.53	1.55	1.50	–71.0
P3(NO ₂)	6-NO ₂ (<i>S</i>)	1.53	1.55	1.50	–56.0
P4(NO ₂)	4-NO ₂ (<i>E</i>)	1.53	1.54	1.50	–58.1
P5(NO ₂)	4-NO ₂ (<i>Z</i>)	1.50	1.55	1.50	–56.8
P6(NO ₂)	1-NO ₂ (<i>S</i>)	1.54	1.54	1.51	–42.7
P7(NO ₂)	1-NO ₂ (<i>R</i>)	1.55	1.54	1.51	–57.0
P8(NO ₂)	5-NO ₂	1.53	1.55	1.50	–59.9
P9(NH ₂)	4-NH ₂ (<i>E</i>)	1.53	1.55	1.50	–53.1
P10(NH ₂)	1-NH ₂ (<i>S</i>)	1.55	1.56	1.50	–62.0
P11(NH ₂)	3-NH ₂ (<i>R</i>)	1.55	1.55	1.50	–54.4
P12(NH ₂)	5-NH ₂	1.53	1.55	1.51	–63.8
P13(NH ₂)	6-NH ₂ (<i>S</i>)	1.53	1.56	1.51	–36.6
P14(NH ₂)	6-NH ₂ (<i>R</i>)	1.53	1.55	1.51	–56.2
P15(NH ₂)	4-NH ₂ (<i>Z</i>)	1.52	1.55	1.51	–61.6

^a Absolute configuration of structures have been shown in the parentheses. For atom numbering see Scheme 1

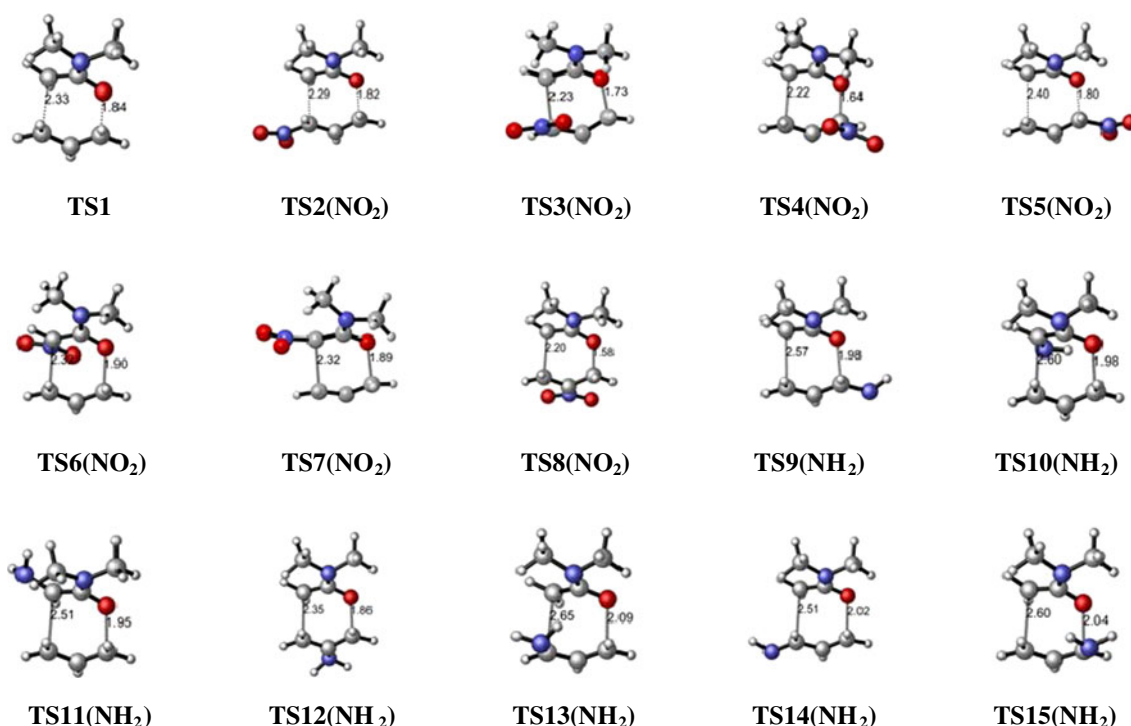
shown in Table 1. As can be seen from this table, dihedral angle of structure **R1** (without any substituent) is 75.2°. In structures **R4(NO₂)**, **R6(NO₂)**, and **R15(NH₂)** the dihedral angles are less than 75.2° (respectively 13.3°, 58.1°, and 53.5°). In structures **R4(NO₂)** and **R15(NH₂)**, the NO₂ and

NH₂ groups with *R* configuration are located in position 4, respectively. In structure **R6(NO₂)**, a NO₂ group with *Z* configuration is located in position 1. In structure **R10(NH₂)**, which a NH₂ group located in the same position with the same configuration with respect to **R6(NO₂)**, the variation of dihedral angle from structure **R1** is only 8.5°. The variation of dihedral angles in structure **R4(NO₂)** may be related to the repulsion between NO₂ group with the large NMe₂ group and the repulsion of the above mentioned group with electron rich π -bonds. Because both of steric and electronic effects in the presence of NO₂ group are more than NH₂ group, therefore, the variation of dihedral angles in structure **R4(NO₂)** is more than structure **R15(NH₂)**. In structure **R6(NO₂)**, the variation of dihedral angle is less than structures **R4(NO₂)** and **R15(NH₂)**. In this structure, there is no repulsion between NO₂ and NMe₂ groups. (More analysis about the structure of reactant, products, and TS structures is available in the [Supplementary material](#).)

Selected bond lengths have been shown in Table 1. Because none of the structures have a ring, therefore substituent effects on the structures are diminished by changing dihedral angles, without appreciable changes in the bond lengths.

Structures of products

In Table 2 selected bond lengths and dihedral angles (C2-C1-C6-C5) and in Fig. 2 the optimized structures of

**Fig. 3** Optimized structures of TSs

products were shown. In structure **P1** the dihedral angle is -55.9 . Structures **P2(NO₂)**, **P8(NO₂)**, **P10(NH₂)**, **P12(NH₂)**, and **P15(NH₂)** have the dihedral angles larger than structure **P1**. In structure **P2(NO₂)**, the increasing of the dihedral angle led to minimum repulsion between NO₂ group with NMe₂ group. In structure **P8(NO₂)**, decreasing of the dihedral angle causes an increase in the repulsion. By decreasing of dihedral angle, steric repulsion between NO₂ and NMe₂ groups and repulsion between electron rich π -bonds will increase. In structure **P10(NH₂)**, any increase of dihedral angle will cause an increase in the electronic repulsion between π -bond with lone-pair electrons of the amine group. On the other hand, any decrease in the dihedral angle will led to the increase in repulsion between π -bonds. In structure **P12(NH₂)**, decrease of the dihedral angle causes an increase of repulsion between π -bonds.

Structures of transition states

In the previous studies on the Claisen rearrangement it had been shown that the chair conformation in TS is more favorable than the boat conformation [28, 29]; therefore chair conformation of TSs has been selected for optimization. Optimized structures of the transition states (TSs) have been shown in Fig. 3. The selected bond length and charges have been presented in Table 3 and Fig. 4, respectively.

Comparison of C1-C6 and O-C4 distances in the TSs clears that this reaction is an asynchronous pericyclic procedure. In all of the cases, the distance between C1-C6, which leads to a creation of a new bond is larger than O-C4. O-C4 is a bond which breaks during the reaction. The distance between C1-C6 varies from 2.20 Å in structure **TS8(NO₂)** to 2.65 Å in structure **TS13(NH₂)**. The distance between O-C4 varies between 1.58 Å in structure **TS8(NO₂)** and 2.09 Å in structure **TS13(NH₂)**. The smallest amount of the difference between C1-C6 and O-C4 is related to structure **TS6(NO₂)** (0.42 Å). This indicates that, this structure has less asynchronous character. The largest difference between the above mentioned distances refers to structures **TS8(NO₂)** and **TS10(NO₂)** (0.62 Å). It means that, these two structures have the most asynchronous character among investigated structures.

The bond order n in the TS structures according to the Pauling relation, $n_p/n_0 = \exp[(d_0 - d_{TS})/C]$, with $n_0=1$ and $C=0.6$ are reported in Table 4 [26]. As can be seen from Table 4 in all of the cases, a high degree of asynchronous bond forming and bond breaking is apparent.

The Mulliken charges of selected atoms which contribute in bond rupture and bond formation in TS have been presented in Fig. 4. In structure **TS1**, all of these atoms (C1, O, C4, C6) are carrying partial negative charges. Introducing a nitro group to the position C6 in structure **TS2(NO₂)**, decreases the negative charge on C6 to -0.01 . On the other hand,

Table 3 Selected distance in the TSs' structure

Entry	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Structure	TS1	TS2 (NO ₂)	TS3 (NO ₂)	TS4 (NO ₂)	TS5 (NO ₂)	TS6 (NO ₂)	TS7 (NO ₂)	TS8 (NO ₂)	TS9 (NH ₂)	TS10 (NH ₂)	TS11 (NH ₂)	TS12 (NH ₂)	TS13 (NH ₂)	TS14 (NH ₂)	TS15 (NH ₂)
C1-C6 (Å)	2.33	2.29	2.23	2.22	2.40	2.32	2.32	2.20	2.57	2.60	2.51	2.35	2.65	2.51	2.60
O-C4 (Å)	1.84	1.82	1.73	1.64	1.80	1.90	1.89	1.58	1.98	1.98	1.95	1.86	2.09	2.02	2.04

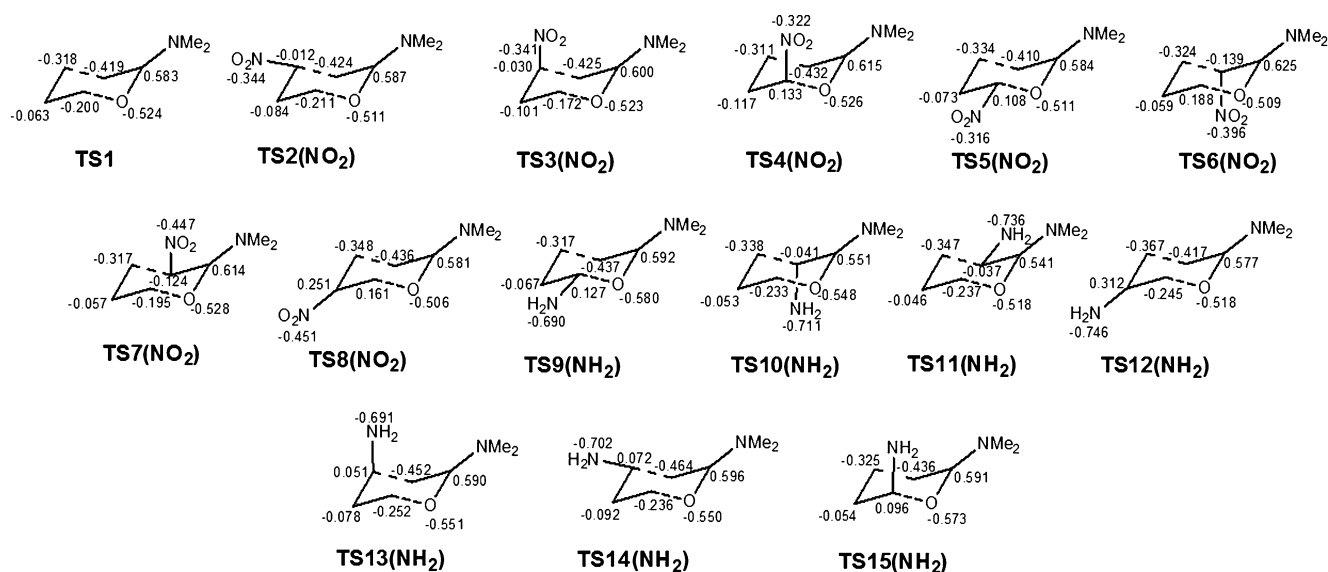


Fig. 4 The Mulliken charge distribution in the TSs' structures

introducing a nitro group to C4 position (structures **TS4(NO₂)** and **TS5(NO₂)**) induces a partial positive charge on C4. As is expected, the stereochemistry of substituent does not alter the atomic charge considerably. Nitro group in position C5 (structure **TS8(NO₂)**) alters the atomic charge of C4 more than C6. The nitro group in the above mentioned structure increases the negative charge on C4 in a small amount.

It could be deduced that because of the existence of resonance between NH₂ group with π -bond in TS, a partial negative charge will be present on the C6 in structures **TS9(NH₂)** and **TS15(NH₂)**. Calculated charges for these structures show that the charge on C6 has not been altered considerably in comparison with structure **TS1**. By investigating dihedral angles between NH₂ group with carbon in these structures (128.3° for structure **TS7(NO₂)** and 127.9° for structure **TS15(NH₂)**), it can be deduced that because of the absence of the co-planarity, there is no resonance between NH₂ group and π -bond in the TS. The comparison between structure **TS4(NO₂)** with **TS15(NH₂)** and **TS5(NO₂)** with **TS9(NH₂)** shows that both of the nitro and amine group have the same effect on the charge distribution in the TS. In all of these structures a partial positive charge has been induced on C4. Between the investigated structures, only structures **TS13(NH₂)** and **TS14(NH₂)** have the partial positive charge on C6 position. It is worthy to mention that in structure **TS2(NO₂)** which contains a nitro group in C6 position, there is a partial negative charge on this center.

Substituent effects on activation energies

In Table 5, the calculated activation energies for the different investigated reactions have been shown. Reported

data shows that the effects of substituent on activation energies depend on the position of substituent more than the nature of substituent. Previously, the effect of electron donating group in Claisen rearrangement has been investigated experimentally [33–37]. These studies showed that the effect of substituent on the accelerating or decelerating of the Claisen rearrangement depends on the position of the substituent. Our results show that the electron withdrawing group in position 5 has the most effect on the activation energy (Table 3, Entry 8). Electron withdrawing groups in this position facilitates the reaction relative to structure **S1** by an amount of 8.6 kcal mol⁻¹. It is worthy to mention that, as mentioned before, structure **S8(NO₂)** has the least separation between O-C4 and C1-C6 bond length in the TS. Electron withdrawing substituent in the position C4 decreases the E_a only 2.8 kcal mol⁻¹. Nitro group in the positions C1, C4, and C6 has little effect on the E_a . The comparison between structures **S4(NO₂)** and **S5(NO₂)** shows that the stereochemistry of the substituent has considerable effect on the E_a (Fig. 5). Structure **S4(NO₂)** has lower E_a relative to structure **S5(NO₂)**.

In spite of the presence of repulsion between NO₂ and NMe₂ in structure **S4(NO₂)**, this structure has lower E_a . The repulsion between NO₂ and NMe₂ does not present in structure **S5(NO₂)**, but this structure has higher E_a . This observation is against the presence of repulsion in the TS structure which should lead to higher E_a .

From Fig. 5, it can be concluded that, the 1,3-diaxial repulsion has the main effect in the reaction procedure. Repulsion between axial NO₂ and NMe₂ group in Fig. 5a, increases the E_a by the amount of 1.7 kcal mol⁻¹. The results presented in Fig. 5b shows the same procedure. In these structures, the NH₂ substituent in the axial position

Table 4 Bond orders calculated according to Pauling relation^a

Entry	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Structure	TS1	TS2	TS3	TS4	TS5	TS6	TS7	TS8	TS9	TS10	TS11	TS12	TS13	TS14	TS15
	(NO ₂)	(NO ₂)	(NO ₂)	(NO ₂)	(NO ₂)	(NO ₂)	(NO ₂)	(NO ₂)	(NH ₂)	(NH ₂)	(NH ₂)	(NH ₂)	(NH ₂)	(NH ₂)	(NH ₂)
O-C4 (%)	51	52	60	67	51	48	48	79	42	40	42	48	34	39	39
C1-C6 (%)	27	29	32	32	24	27	27	33	18	17	20	27	16	20	17

^a based on the equation: $n_o = 1; C = 0.6; d_o = \text{distance O-C4 or C1-C6 in the reactant and product, respectively}; d_{TS} = \text{distance O-C4 or C1-C6 in the TS structure}$

Table 5 Effect of substitution and its position on ΔE_a

Entry	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Structure	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
	(NO ₂)	(NO ₂)	(NO ₂)	(NO ₂)	(NO ₂)	(NO ₂)	(NO ₂)	(NO ₂)	(NO ₂)	(NH ₂)	(NH ₂)	(NH ₂)	(NH ₂)	(NH ₂)	(NH ₂)
ΔE_a (kcal/mol)	19.5	19.0	20.7	16.6	19.6	20.8	23.6	10.9	12.6	18.6	18.8	23.8	18.3	17.5	15.2
$\Delta\Delta E_a$	0.0	-0.5	1.2	-2.8	0.1	1.3	4.1	-8.6	-6.9	-0.9	-0.7	4.3	-1.2	-2.0	-4.2

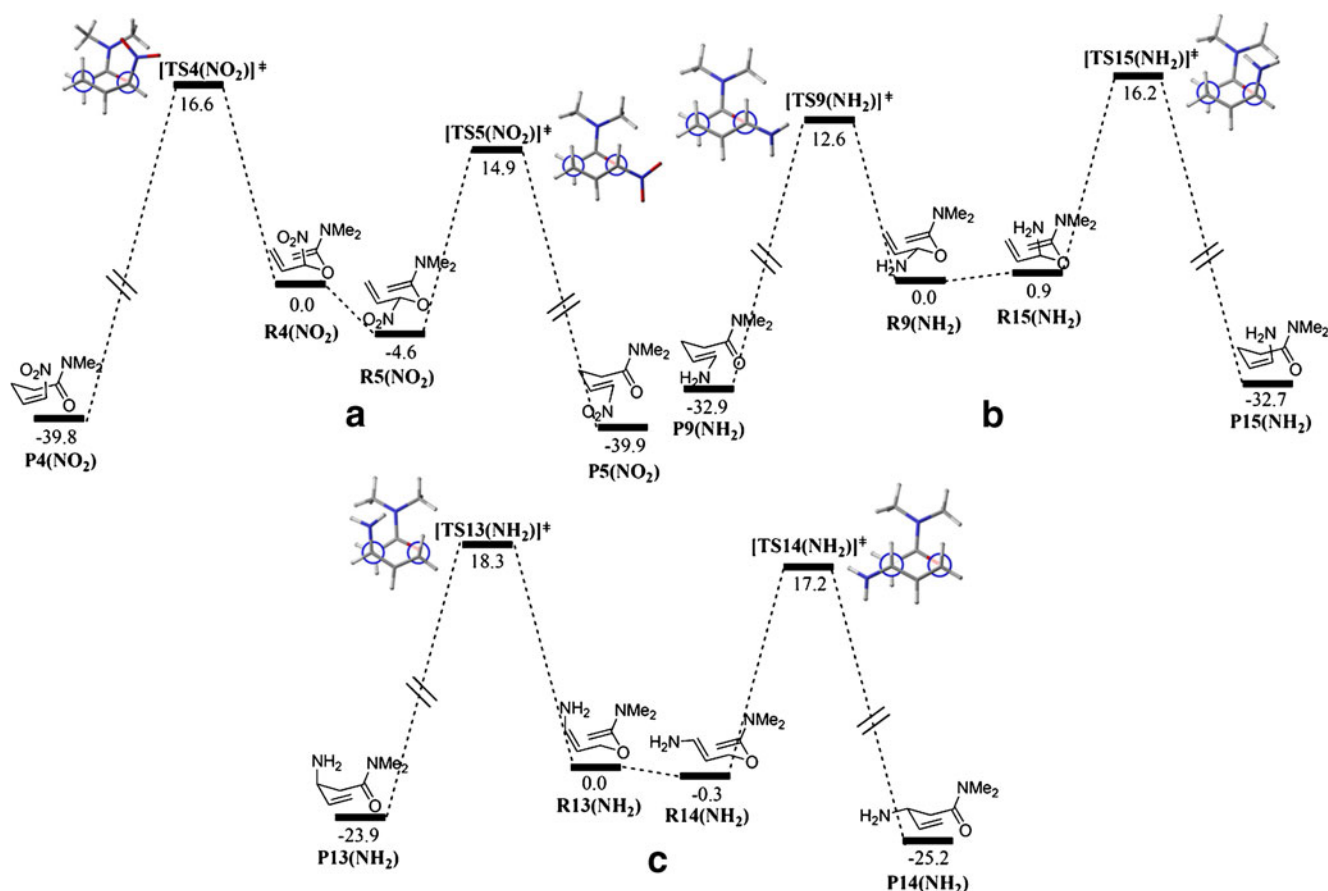


Fig. 5 Energy profiles of pairs which have different stereochemistry of one carbon

increases the E_a by the amount of $2.6 \text{ kcal mol}^{-1}$ in comparison with equatorial substituent.

By comparison of **a** and **b**, it could be concluded that, because NO_2 group is larger than NH_2 group, the 1,3-diaxial repulsion in structure **a** should have more effect than **b**. The obtained results are in opposite direction. This means that, 1,3-diaxial repulsion is one of the reasons for the observed increase in the E_a , but it is not the sole reason. By comparison of the charges of the structures which are presented in the Fig. 4, another reason could be mentioned. The amount of the change in the charges in structure **TS4**(NO_2) and **TS5**(NO_2) is little on the C-4. Therefore the difference in E_a of these two structures is little. The comparison of structures **TS9**(NH_2) with **TS15**(NH_2) shows that, the changes of the partial charges in these two structures is larger on C-4. The positive charge on the C-4 will cause a decrease in the E_a . It can be attributed to the greater attraction between positive C-4 with negative O in the TS structure which leads to the decrease in the E_a .

The same kind of interpretation can be done for the Fig. 3c. In these two structures, the NH_2 group in the axial position has higher activation energy; but its amount is only $1.1 \text{ kcal mol}^{-1}$. It means that, 1,3-diaxial repulsion is not

the main factor. The evaluation of the charges on the C6 in these two structures reveals that, when the NH_2 group is in axial position, the partial positive charge on C6 is larger. This will cause an increased attraction between positive charge on C6 with partial negative charge on C1.

The higher E_a of structure **S12**(NH_2) in contrast to **S1**, can be interpreted by similar way. In this structure, NH_2 and NMe_2 groups are far from each other and repulsion between these two groups diminishes to the least amount. By considering the charges on the C4 and C6 in these structures, it can be concluded that the amount of negative charge on these two atoms in the **S12**(NH_2) is larger than **S1**. This will cause the greater repulsion between charges and an increase in the E_a .

The comparison of structures **S2**(NO_2) and **S3**(NO_2) shows that the effect of the NO_2 group on the charge distribution on the C4 is greater than C6. It is worthy to mention that, in these two structures, the NO_2 group is positioned on C-6 not C4. Because of non-planarity of the NO_2 group with double-bond in the TS structure, electron withdrawing effect with mesomeric effect can be neglected. The larger negative charge on C-4 in structure **S2**(NO_2) will increase the E_a ; but in structure **S3**(NO_2), the 1,3-diaxial

repulsion will increase the E_a . Because these two effects are playing the main role in structures **S2(NO₂)** and **S3(NO₂)** respectively, the E_a of these two structures are very similar. The comparison of these two structures shows that the effect of 1,3-diaxial repulsion is more important than the other one.

Conclusions

In conclusion, we have investigated substituent effect on the Meerwein–Eschenmoser–Claisen rearrangement. The results showed that the substituent effect not only depends on the nature of substituent, but also on its position and configuration. The nitro group in position 5 and the amine group in positions 4 (with *S* and *R* configuration respectively) have the most accelerating effect on the EMC-R. The effect of amine group in position 5 is decelerating. Substituent effects on other positions are negligible. In some cases the 1,3-diaxial repulsion has the main effect in the reaction procedure; in other ones, 1,3-diaxial repulsion plays a minor role. In these cases the partial charges on the atoms and attraction or repulsion between them in TS structures could be used to analyze the observed changes in E_a . As an example structure **S12(NH₂)** could not be analyzed by repulsion between substituents; but the results are interoperable by considering the partial charges. In this investigation, only the substituents which have mesomeric effect have been investigated and the results could be different with substituents having an inductive effect only. (Cartesian coordinates and energies of optimized structures were reported in the supporting information.)

Acknowledgments Special thanks to Dr. Shant Shahbazian for his assistance and guidance during the study and preparing of this manuscript. We thank Professor Seik Weng Ng for making available to us his softwares and machine time facilities. The financial support of Research Council of Shahid Beheshti University is gratefully acknowledged.

References

1. Castro AMM (2004) Claisen rearrangement over the past nine decades. *Chem Rev* 104:2939–3002
2. Meerwein H, Stopp G, Florian W, Schon N (1961) *Liebigs Ann Chem* 641:1–39
3. Wick AE, Steen K, Felix D, Eschenmoser A (1964) *Helv Chim Acta* 47:2425–2429
4. Hiersemann M, Nubbemeyer U (2007) *The claisen rearrangement: methods and applications*. Wiley-VCH, Weinheim
5. Wovkulich PM, Tang PC, Chadha NK, Batcho AD, Barrish JC, Uskoković MR (1989) Remote diastereoselection in the asymmetric total synthesis of mevinolin. *J Am Chem Soc* 111:2596–2599
6. Yoshida M, Shoji Y, Shishido K (2009) Total syntheses of enokipodins a and b utilizing palladium-catalyzed addition of an arylboronic acid to an allene. *Org Lett* 11:1441–1443
7. Majewski M, Snieckus V (1984) Synthesis of pyrethroid amides via epoxy amide cyclization. *J Org Chem* 49:2682–2687
8. Tulshian DB, Fraser-Reid B (1984) Routes to c-glycopyranosides via sigmatropic rearrangements. *J Org Chem* 49:518–522
9. Qu H, Gu X, Min BJ, Liu Z, Hrubby VJ (2006) Synthesis of anti- β -substituted γ , δ -unsaturated amino acids via eschenmoser-claisen rearrangement. *Org Lett* 8:4215–4218
10. Tsang R, Fraser-Reid B (1985) A route to optically active trichothecane skeleton by bisannulation of a pyranose derivative. *J Org Chem* 50:4659–4661
11. Gilbert MW, Galkina A, Mulzer J (2004) Toward the total syntheses of pepluanin a and euphosalicin: concise route to a highly oxygenated cyclopentane as a common intermediate. *Synlett* 14:2558–2562
12. Loh TP, Hu QY (2001) Synthetic studies toward anisatin: a formal synthesis of (\pm)-8-deoxyanisatin. *Org Lett* 3:279–281
13. Gradl SN, Kennedy-Smith JJ, Kim J, Trauner D (2002) A practical variant of the claisen-eschenmoser rearrangement: synthesis of unsaturated morpholine amides. *Synlett* 3:411–414
14. Chen CY, Hart DJ (1990) Total synthesis of dl-stenine. *J Org Chem* 55:6236–6240
15. Kündig EP, Laxmisha MS, Cannas R, Tchertchian S, Ronggang L (2005) Chromium-mediated dearomatization: application to the synthesis of racemic 15-acetoxytubipofuran and asymmetric synthesis of both enantiomers. *Helv Chim Acta* 88:1063–1080
16. Linton EC, Kozlowski MC (2008) Catalytic enantioselective meerwein-eschenmoser claisen rearrangement: asymmetric synthesis of allyl oxindoles. *J Am Chem Soc* 130:16162–16163
17. Rao Lingam VSP, Vinodkumar R, Mukkanti K, Thomas A, Gopalan B (2009) Tetra-n-butylammonium fluoride-catalyzed eschenmoser-claisen [3,3]-sigmatropic rearrangement. *Synth Commun* 39:332–341
18. Burrows CJ, Carpenter BK (1981) Substituent effects on the aliphatic claisen rearrangement. I. Synthesis and rearrangement of cyano-substituted allyl vinyl ethers. *J Am Chem Soc* 103:6983–6984
19. Cooper JA, Olivares CM, Sandford G (2001) Nucleophilic substitution and claisen rearrangement reactions of model fluoroalkenes of general structure $r\text{-cf}=\text{cf}\text{-cf}₃$. *J Org Chem* 66:4887–4891
20. Davidson MM, Hillier IH, Vincent MA (1995) The claisen rearrangement of allyl vinyl ether in the gas phase and aqueous solution. Structures and energies predicted by high-level ab initio calculations. *Chem Phys Lett* 246:536–540
21. Dewar MJS, Healy EF (1984) Mndo study of the claisen rearrangement. *J Am Chem Soc* 106:7127–7131
22. Dewar MJS, Jie C (1989) Mechanism of the claisen rearrangement of allyl vinyl ethers. *J Am Chem Soc* 111:511–519
23. Gajewski JJ, Gee KR, Jurayj J (1990) Energetic and rate effects of the trifluoromethyl group at c-2 and c-4 on the aliphatic claisen rearrangement. *J Org Chem* 55:1813–1822
24. Khaledy MM, Kalani MYS, Khuong KS, Houk KN, Aviyente V, Neier R, Soldermann N, Velker J (2003) Origins of boat or chair preferences in the ireland-claisen rearrangements of cyclohexenyl esters: a theoretical study. *J Org Chem* 68:572–577
25. McMichael KD, Korver GL (1979) Secondary deuterium isotope effects and transition state structure in the aromatic claisen rearrangement [12]. *J Am Chem Soc* 101:2746–2747
26. Meyer MP, DelMonte AJ, Singleton DA (1999) Reinvestigation of the isotope effects for the claisen and aromatic claisen rearrangements: the nature of the claisen transition states. *J Am Chem Soc* 121:10865–10874

27. Schuler FW, Murphy GW (1950) The kinetics of the rearrangement of vinyl allyl ether. *J Am Chem Soc* 72:3155–3159
28. Vance RL, Rondan NG, Houk KN, Jensen F, Borden WT, Komornicki A, Wimmer E (1988) Transition structures for the claisen rearrangement. *J Am Chem Soc* 110:2314–2315
29. Wiest O, Black KA, Houk KN (1994) Density functional theory isotope effects and activation energies for the cope and claisen rearrangements. *J Am Chem Soc* 116:10336–10337
30. Frisch MJ, Trucks GW, Schlegel HB, Gill PMW, Johnson BG, Robb MA, Cheeseman JR, Keith T, Petersson GA, Montgomery JA, Raghavachari K, Al-Laham MA, Zakrzewski VG, Ortiz JV, Foresman JB, Cioslowski J, Stefanov BB, Nanayakkara A, Challacombe M, Peng CY, Ayala PY, Chen W, Wong MW, Andres JL, Replogle ES, Gomperts R, Martin RL, Fox DJ, Binkley JS, Defrees DJ, Baker J, Stewart JP, Head-Gordon M, Gonzalez C, Pople JA (1998) Gaussian 98, Revision A.9. Gaussian Inc, Pittsburgh
31. Frisch A, Nielsen AB, Holder AJ (2000) Gaussview users manual. Gaussian Inc, Pittsburgh
32. CYLview bL, CY, Université de Sherbrooke, 2009 (<http://www.cylview.org>)
33. Koreeda M, Luengo JI (1985) Anionic oxy-claisen rearrangement of enolates of α -allyloxy ketones. A remarkable rate-accelerating effect exhibited by the nature of the counterion. *J Am Chem Soc* 107:5572–5573
34. Barluenga J, Aznar F, Liz R, Bayod M (1984) Synthesis of substituted 2-aminopent-4-enals and 2-amino-3-(2-furyl) propenals via [3,3]- and [1,3]-sigmatropic shifts of β -allyloxyenamines. *J Chem Soc Chem Commun* 21:1427–1428
35. Welch JT, Samartino JS (1985) Facile diastereoselective ester enolate claisen rearrangements of allyl fluoroacetates. *J Org Chem* 50:3663–3665
36. Frey HM, Montague DC (1968) Thermal unimolecular isomerization of 1-methylallyl vinyl ether in the gas phase and in solution. *Trans Faraday Soc* 64:2369–2374
37. Wilcox CS, Babston RE (1986) Substituent effects in [3,3]-sigmatropic rearrangements. Alkyl group effects and transition-state "Syn-diaxial" interactions. *J Am Chem Soc* 108:6636–6642

A DFT study on equilibrium geometries, stabilities, and electronic properties of small bimetallic Na-doped Au_n ($n=1-9$) clusters: comparison with pure gold clusters

Yan-Fang Li · Xiao-Yu Kuang · Ai-Jie Mao · Yang Li ·
Ya-Ru Zhao

Received: 18 February 2011 / Accepted: 24 March 2011 / Published online: 27 April 2011
© Springer-Verlag 2011

Abstract A systematic study on the geometric structures, relative stabilities, and electronic properties of small bimetallic Au_nNa ($n=1-9$) clusters has been performed by means of first-principle density functional theory calculations at the PW91PW91 level. The results show that the optimized ground-state isomers adopt planar structures up to $n=5$, and the Na-capped geometries are dominant growth patterns for $n=6-9$. Dramatic odd-even alternative behaviors are obtained in the second-order difference of energies, fragmentation energies, highest occupied-lowest unoccupied molecular orbital energy gaps, and chemical hardness for both Au_nNa and Au_{n+1} clusters. It is found that Au_5Na and Au_6 have the most enhanced stability. Here, the size evolutions of the theoretical ionization potentials are in agreement with available experimental data, suggesting a good prediction of the lowest energy structures in the present study. In addition, the charge transfer has been analyzed on the basis of natural population analysis.

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1073-9) contains supplementary material, which is available to authorized users.

Y.-F. Li · X.-Y. Kuang (✉) · A.-J. Mao · Y.-R. Zhao
Institute of Atomic and Molecular Physics, Sichuan University,
Chengdu 610065, China
e-mail: scu_kuang@163.com

Y.-F. Li
e-mail: scu_lyf@163.com

X.-Y. Kuang
International Centre for Materials Physics, Academia Sinica,
Shenyang 110016, China

Y. Li
Department of Opto-Electronics Science and Technology,
Sichuan University,
Chengdu 610065, China

Keywords Au-Na cluster · Density functional theory ·
Geometric configuration · Ionization potentials

Introduction

Clusters are aggregates containing from a few to many millions of atoms or molecules. There is continuing interest in metal clusters, in part, because of their potential applications in fields such as catalysis, optics, magnetics, nano-electronics, and nanostructured materials [1–13]. As for alkali or noble metals, since the closed subshell and limited number of valence electrons can greatly simplify theoretical simulation and spectroscopic measure, they are particularly convenient for investigations. With this in mind, numerous works have presented an in-depth study on chemical and physical properties of finite clusters [14–23]. Interesting, it is found that the alkali-metal clusters, with the size up to thousands of atoms, confirm to the jellium model. In jellium model, the magic-sized clusters are relatively stable due to their filled electronic shells. Contrarily, this feature is suitable for the small-sized clusters of noble metals [24, 25].

Heteronuclear clusters are particularly interesting as they may display structures and properties which are distinct from those of the corresponding pure nanoparticles. The mixed metal dimers, serving as the combinations of an alkali atom and a noble-metal atom, have been widely explored to understand the microscopic details of heteronuclear metal-metal bond, such as NaAu, NaAg, CsAu, LiAg, LiCu, KAg, and so on [26–31]. Among them, Stangassinger et al. [31] have studied metal diatomics KAu and NaAu by using resonant two-photon ionization electronic spectroscopy. For both heteronuclear diatomics, the dissociation energies are provided, and it is found that they are significantly greater than those of the

corresponding homonuclear diatomics Na_2 , K_2 , or Au_2 . Moreover, the complete alkali auride series (LiAu , NaAu , KAu , RbAu , CsAu) have been studied by Belpassi et al. [26] in the program BERTHA. The electronic properties change dramatically along these alkali metal series, and the intermetallic bond that occurs in them is highly polar and is characterized by a large charge transfer from the alkali metal to gold atom. Apart from the above mentions, since 1986, single-crystal data for NaAu , RbAu , and CsAu have been available [32, 33]; subsequently, Watson et al. [34] showed that the semiconducting gap in solid CsAu and RbAu originates from a combination of the Au–Au separation and the ionic character of the compounds; also, in order to search for ideal plasmonic materials, the electronic structure calculations have been performed among the alkali-noble diatomics [29]. In recent years, more and more attention has been focused on the size-dependent evolutions because the structures and properties of bimetallic clusters (A_mB_n) may be tuned by varying the composition. Therefore, the relative thermodynamic stability, ionization potential, and bonding character of Au_mNa_n , Ag_mNa_n , and Au_mCs_n have been experimentally investigated [35–37]. On the theoretical side, although the intermetallic molecules NaAu , AuNa_n , Au_5Na , and Au_{19}Na have been studied by employing density functional theory (DFT) previously [22, 38–40], few systematically theoretical investigations on the bimetallic Au_nNa clusters have been performed. Are their structures and properties greatly distinct from the bare clusters? How does the composition of dopant affect the growth pattern? These open questions are still unsolved. Thus, in order to provide further insight on Au_nNa clusters, we have reported a DFT investigation on the small Au_nNa ($n=1-9$) clusters. The geometric structures, relative stabilities, natural population analysis (NPA), ionization potentials and chemical hardness have been analyzed. This paper is organized as: “Computational details” section gives a brief description of the theoretical approach. Then, detailed results and discussions are presented in “Results and discussions” section. Lastly, the conclusions are summarized in “Conclusions” section.

Computational details

In this work, the geometry optimizations on Au_nNa ($n=1-9$) clusters are performed by using a DFT-based *GAUSSIAN 03* package [41]. Because the gold is a heavy atom, calculations including all electrons are out of consideration and the effective core potential (ECP) are introduced. Thus, the Au atom is treated with the Stuttgart/Dresden double- ζ SDD (ECP60MWB) basis set [42, 43]. The 6-311G* basis set is employed for all electrons in Na. In addition, literature reports on gold-sodium system have

emphasized the feasibility of gradient-corrected exchange and correlation functional *Perdew-Wang* (PW91PW91) [44]. Then, all the calculations, such as geometry optimization, highest occupied-lowest unoccupied molecular orbital (HOMO-LUMO) energy gap, vertical ionization potential (VIP), vertical electron affinity (VEA) as well as NPA, are conducted at the PW91PW91/GEN (SDD for the Au atom and 6-311G* for the Na atom) method level. Here, the configuration is regarded as optimized when the maximum force, root-mean-square (RMS) force, maximum displacement of atoms, and RMS displacement of atoms have magnitudes less than 0.00045, 0.0003, 0.0018, and 0.0012 au, respectively. Considering the spin polarization, different spin multiplicities are also taken into account for every initial configuration. Besides, all geometries have been fully relaxed and frequency analysis has been carried out for the purpose of determining the structures to correspond to the local minima in the energy surface. Since a large number of configurations have been optimized, we focus the particular attention only on the energetically low-lying ones. The other high-energy configurations are shown in electronic supplementary material (ESM).

Results and discussions

Equilibrium geometries

For the Au_nNa ($n=1-9$) clusters, the geometric structures of the lowest energy configurations and low-lying isomers are presented in Fig. 1. Meanwhile, the detailed information for their structural characteristics (symmetry, electronic state, relative energy, HOMO and LUMO energies, and the lowest vibrational frequency) is given in Table 1. In order to discuss the effects of doped impurity atom on gold clusters, we first perform the calculations on Au_{n+1} ($n=1-9$) clusters, and the lowest energy structures are also shown in Fig. 1. All the Au_{n+1} structures are in line with the previous DFT studies, and the spin states are found to prefer the lowest multiplicity.

The heteronuclear diatomic AuNa has a bond length of 2.617 Å, and the electronic state is singlet ($^1\Sigma$). The calculated dissociation energy of Au–Na dimer is 2.55 eV, which is larger than that of Au_2 (2.30 eV); but it is in good agreement with available experimental results (2.64±0.2 [31], 2.697 [35], and 2.2±0.13 eV [45]).

As for small-sized Au_nNa cluster, the geometry of Au_2Na in the lowest energy state 2a is an isosceles triangle possessing two identical Au–Na bonds (2.770 Å). Its apex angle is 58.7°. The low-lying isomer 2b is a linear structure with the Na atom in the center. Considering the different spin multiplicities, all possible structures of Au_3Na clusters are optimized. The optimized isomers 3a and 3b can be viewed as one gold atom adding to the different positions of

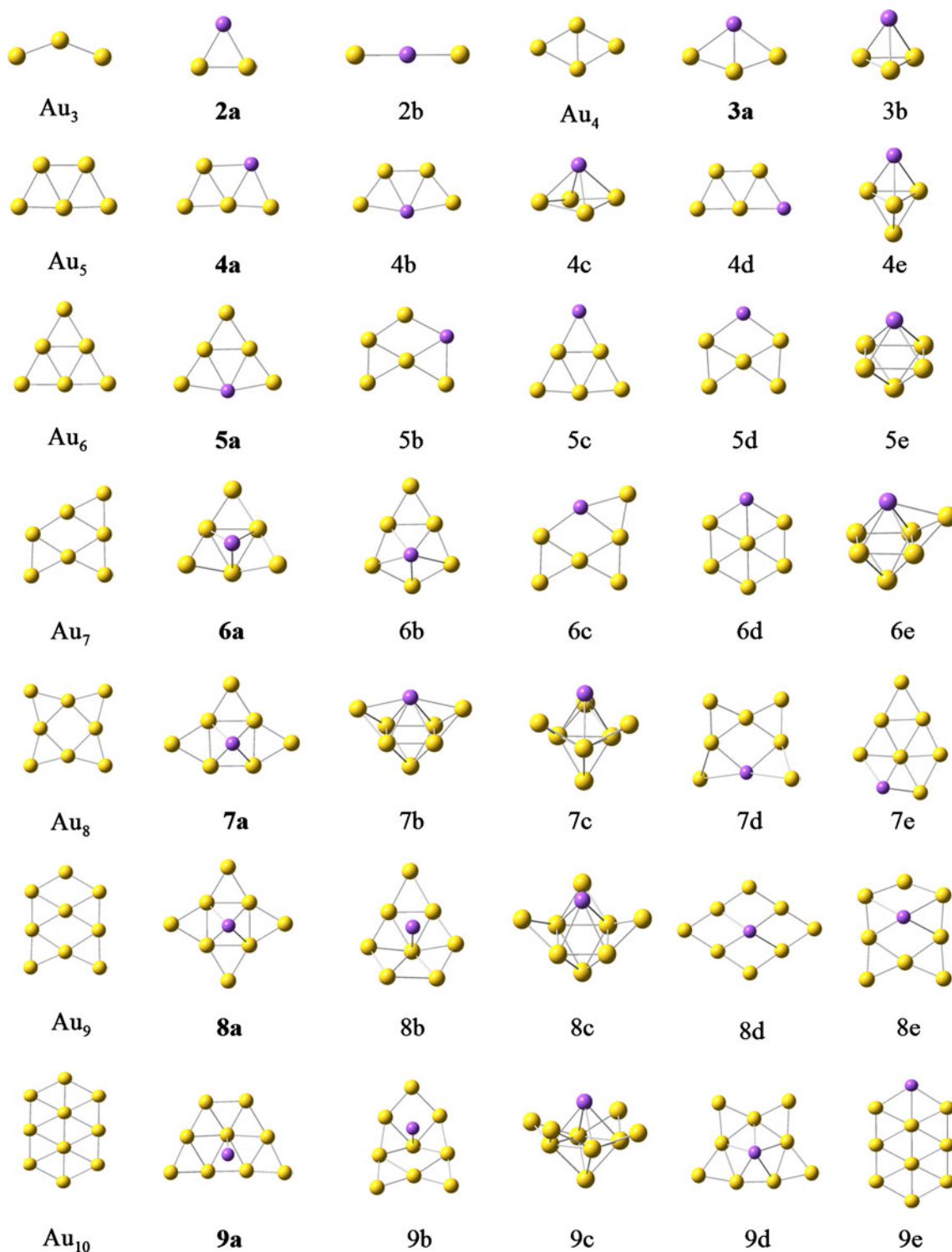


Fig. 1 The lowest energy structures and low-lying isomers for Au_nNa ($n=1-9$) clusters, and the ground-state structures of pure gold clusters Au_{n+1} ($n=1-9$) have been listed on the left. The yellow and purple balls represent Au and Na atoms, respectively

isomer 2a. The ground-state structure 3a is a plane C_{2v} rhombus with a singlet electronic state (1A_1), while that of the 3b is three-dimensional (3D) with a higher electronic state 3A_1 . It is worth pointing out that 3b is the first appearance of 3D geometry for Au_nNa ($n=1-9$) clusters.

For Au_2Na and Au_3Na clusters, the energy differences between the lowest energy configurations and the isomers are 0.75 and 1.09 eV, respectively. Here, our optimized lowest energy structures of pure gold clusters prefer an obtuse triangle and rhombus, respectively.

Table 1 Symmetries, electronic states, relative energies (ΔE), HOMO energies, LUMO energies, and the lowest vibrational frequencies for Au_nNa ($n=1-9$) clusters. The lowest energy isomers are denoted in bold

Isomer	Symm.	State	ΔE (eV)	HOMO (hartree)	LUMO (hartree)	Freq. (cm^{-1})
1a	$C_{\infty v}$	$^1\Sigma$	0.00	-0.16964	-0.08619	236.9
2a	C_{2v}	2B_2	0.00	-0.15121	-0.13022	117.9
2b	D_{3h}	$^2\Sigma_u$	0.75	-0.21237	-0.20342	27.2
3a	C_{2v}	1A_1	0.00	-0.19594	-0.12312	39.6
3b	C_{3v}	3A_1	1.09	-0.16519	-0.14342	73.7
4a	C_s	$^2A'$	0.00	-0.19722	-0.18021	20.7
4b	C_{2v}	2A_1	0.06	-0.19230	-0.17962	16.2
4c	C_{2v}	2B_1	0.13	-0.16917	-0.15609	35.9
4d	C_s	$^2A'$	0.47	-0.16336	-0.14993	17.7
4e	C_{3v}	4A_1	1.27	-0.17945	-0.16196	50.3
5a	C_{2v}	1A_1	0.00	-0.21566	-0.13673	29.9
5b	C_s	$^1A'$	0.51	-0.20505	-0.14288	4.7
5c	C_{2v}	1A_1	0.58	-0.18646	-0.12160	16.7
5d	C_{2v}	1A_1	0.92	-0.18363	-0.14768	23.3
5e	C_{4v}	1A_1	1.22	-0.18856	-0.16110	7.4
6a	C_s	$^2A'$	0.00	-0.15564	-0.14545	20.6
6b	C_s	$^2A'$	0.04	-0.17010	-0.16092	11.8
6c	C_s	$^2A'$	0.17	-0.17985	-0.16656	9.2
6d	C_s	$^2A''$	0.18	-0.16267	-0.15313	8.5
6e	C_s	$^2A'$	0.22	-0.18806	-0.17789	21.2
7a	C_s	$^1A'$	0.00	-0.18786	-0.14263	10.1
7b	C_{2v}	1A_1	0.10	-0.19918	-0.14276	28.5
7c	C_{3v}	1A_1	0.14	-0.19689	-0.13683	22.1
7d	C_2	1A	0.18	-0.21966	-0.15949	2.3
7e	C_I	1A	0.30	-0.18159	-0.14312	11.8
8a	C_{4v}	2A_1	0.00	-0.16627	-0.15892	7.6
8b	C_I	2A	0.28	-0.17591	-0.16799	7.1
8c	C_s	$^2A'$	0.35	-0.19171	-0.18355	10.4
8d	C_{2v}	2A_2	0.46	-0.17598	-0.16950	10.1
8e	C_s	$^2A'$	0.46	-0.19917	-0.19225	7.1
9a	C_I	1A	0.00	-0.20011	-0.15744	8.1
9b	C_s	$^1A'$	0.04	-0.21475	-0.16439	9.8
9c	C_{2v}	1A_1	0.05	-0.20158	-0.13770	4.4
9d	C_s	$^1A'$	0.15	-0.20010	-0.16027	6.9
9e	C_s	$^1A'$	0.24	-0.19062	-0.15138	3.8

Five possible geometries of Au_4Na cluster, which are described as the trapezoidal configurations 4a, 4b, and 4d, the deformed-square pyramid 4c, and the trigonal bipyramid 4e, are optimized, together with the harmonic frequency analysis. As shown in Fig. 1, with the impurity substituting the different positions of the ground-state structure of Au_5 cluster, the isomers 4a, 4b, and 4d are similar to each other. Interestingly, the low-lying structure 4b in a high coordinated position and geometric symmetry (C_{2v}) is not corresponding to the lowest energy; whereas it is almost energetically degenerate for isomer 4a (0.06 eV). The electronic state and symmetry in both isomers 4a and 4d is $^2A'$ and C_s , while the relative energies are 0.00 and 0.47 eV, respectively. Due to the Jahn-Teller

effect, the 3D isomer 4c is somewhat deformed from square pyramid. Consequently, the symmetry is lowered from C_{4v} to C_{2v} . As for the least stable 3D structure 4e, it can be yielded by capping one Au atom on the bottom side of Au_3Na (3b). This configuration, with three identical Au-Na bonds (2.909 Å), is also in a high electronic state (4A_1).

The lowest energy structure 5a of Au_nNa cluster with six atoms is also found to adopt planar form, which can be described as the impurity being located centrally on one side of the equilateral triangle Au_6 . Thus, the D_{3h} symmetry distorts to C_{2v} . When the sodium atom sits in the apex position of the equilateral triangle, the isomer 5c is born. It is also C_{2v} symmetry and 1A_1 electronic state, but it is 0.58 eV higher in total energy than isomer 5a. The planar

isomers 5b and 5d have a similar structure in which the sodium atom occupies different locations. The different relative energies (0.51 eV, 0.92 eV), symmetries (C_s , C_{2v}), and electronic states ($^1A'$, 1A_1) exist between them. As for the 3D structure, isomer 5e, with four equal Au–Na bonds (2.914 Å), is the least stable configuration (1.22 eV). It is noticed that the structures of Au_5Na for 5a, 5b, and 5c reported by Majumder et al. [39] match with our calculated results.

For the possible Au_nNa isomers with the lowest energy, the calculated results indicate that the 2D→3D structural transition occurs at $n=6$. The 6a configuration, which is obtained after one Na atom is capped on the ground state of Au_6 , is optimized to be the most stable structure. Because the adding impurity Na slightly deforms the initial coordinates, in which we find two different sets of direct Au–Na bonds, those are, 2.922 Å and 3.058 Å. Meanwhile, the other 3D isomer 6b is obtained in the optimizing progress, which also can be viewed as the sodium atom occupying a capping position. The energy difference between these two isomers are small (0.04 eV) and the structures can be said to be energetically degenerate at the present PW91PW91/GEN level. The geometric symmetries and electronic states of them are C_s and $^2A'$, respectively. With the impurity substituting the different positions of the ground-state structure of Au_7 cluster, surprisingly, it is found that the structure with the Na atom occupying a peripheral position (6c) is more preferred than that of central position. This phenomena also can be verified in the other isomer 6d. The deformed 3D hexagon 6d has two different Au–Na bonds, which are 2.883 and 2.971 Å. Its corresponding electronic state is $^2A''$, and the relative energy is 0.18 eV as compared with 6a. Besides, with one gold atom being surface capped on one face of isomer 5e, the configuration 6e can be yielded.

With regard to Au_7Na cluster, the calculation again turns out 3D structure (7a) to be the lowest in energy. This configuration, with C_s symmetry, can be described as the ground-state structure of Au_7 being capped by a sodium atom. Its electronic state prefers the lowest multiplicity ($^1A'$). The other four low-lying isomers (7b, 7c, 7d, 7e) are also 3D structures, which geometric symmetries and electronic states, respectively, are C_{2v} , C_{3v} , C_2 , C_1 and 1A_1 , 1A_1 , 1A , 1A . Also, the corresponding relative energies are 0.10, 0.14, 0.18 as well as 0.30 eV. One can find that the second-stable structure, isomer 7b, can be viewed as two Au atoms bi-capped on the 5e isomer. In this way, the isomer 7c is generated by adding three golds to trigonal bipyramid 4e. So the isomer 7c has a high geometric symmetry (C_{3v}); however, the calculation indicates that it is not the ground-state configuration. Here, the addition of gold atoms makes the Au–Na bonds vary from 2.909 Å (4e) to 3.072 Å (7c). The 7d isomer, in which two bottom gold atoms are twisted

in opposite directions, is born after one Au atom being substituted by sodium in the ground state Au_8 . In structure 7e, we also find that the sodium atom prefers a peripheral position.

Guided by the predicted growth pattern on the Au_nNa clusters as above, five 3D isomers at $n=8$, emerging in the energy range 0.46 eV, are obtained and shown in Fig. 1. From the findings related to the ground-state geometry, the 8a isomer is generated with the sodium atom capped on the Au_8 frame. This structure is in high symmetry, with four equal Au–Na bonds (2.959 Å), and its electronic state is 2A_1 . The least stable isomer 8e resembles the ground-state Au_9 , whereas the Na atom is surface capped. This consists of the optimized result of isomer 8d, in which the Na atom is capped on a rhombus. Here, the less stable isomer 8b is optimized by bonding the impurity atom to the central gold of the second-stable structure of Au_8 in 2.784 Å. Its symmetry is lowered to C_1 . According to the initial geometry of 7b, the other low-lying configuration 8c is obtained, whose energy is 0.35 eV higher than that of isomer 8a.

For Au_9Na clusters shown in Fig. 1, five singlet stable structures, labeled as 9a, 9b, 9c, 9d, and 9e, are optimized to be the minima at the energy surface. All of them again correspond to 3D geometry. On the basis of trapezoidal Au_9 configuration, the most stable 9a structure can be formed when the central Au atom is capped by one Na atom. The Au–Na bond length is 2.827 Å, whose symmetry and electronic state, respectively, are C_1 and 1A . Similar to the Au_nNa isomers 6a, 7a, and 8a, one Na-capped Au_9 structure 9b is considered. Although the geometric characteristics of 9a and 9b are highly different, their total energies are almost degenerate by 0.04 eV. The Au–Na bond in 9b isomer is 3.038 Å. With a relative higher symmetry (C_{2v}), the isomer 9c is optimized as one gold atom surface capped on configuration 8c, whose electronic state is 1A_1 . Also, isomer 9c is degenerate with 9a and 9b in total energy. Here, a new isomer 9d with Na atom being centrally localized at the capping position is also a stable structure. It looks like a butterfly, and its stability is weaker as compared to the 9a isomer by 0.15 eV. Different from the planar structure of ground-state Au_{10} cluster, the least stable isomer 9e has a slight deformation in order to reach the minimum of the potential surface. Three Au–Na bonds relating to the apical impurity are 2.904, 2.969, and 2.904 Å, respectively. For both isomers 9d and 9e, the symmetries and electronic states are identical, namely, C_s and $^1A'$.

An overview of all optimized structures shows that in Au_nNa clusters the Na atom prefers a peripheral position, and the Na-capped Au_n clusters become the dominant geometries for the larger clusters. This may be due to the fact that the Na atom has a larger ionic radius (1.570 Å) than that of Au atom (1.370 Å). In addition, for the lowest

energy states, the Na atom corresponds to a planar configuration for $n \leq 5$, while Au_nNa cluster favors 3D structure from $n=6$ up to $n=9$. In other words, the geometric transition from 2D to 3D occurs at the cluster size $n=6$.

Relative stability

Based on the lowest energy structures, we now discuss the relative stabilities of Au_nNa and Au_{n+1} ($n=1-9$) clusters by examining the atomic average binding energies (E_b), which are expressed as

$$\begin{aligned} E_b(\text{Au}_n\text{Na}) &= [nE(\text{Au}) + E(\text{Na}) - E(\text{Au}_n\text{Na})]/(n+1), \\ E_b(\text{Au}_{n+1}) &= [(n+1)E(\text{Au}) - E(\text{Au}_{n+1})]/(n+1), \end{aligned} \quad (1)$$

where $E(\text{Na})$, $E(\text{Au})$, $E(\text{Au}_n\text{Na})$, and $E(\text{Au}_{n+1})$, respectively, are the total energies of the ground-state Na, Au, Au_nNa ,

and Au_{n+1} . In cluster physics, the second-order difference of energies ($\Delta_2 E$), which can be compared with the relative abundances determined in mass spectroscopy experiment, are a quite sensitive quantity that reflects the relative stability of cluster. For Au_nNa and Au_{n+1} clusters, the corresponding definitions are

$$\begin{aligned} \Delta_2 E(\text{Au}_n\text{Na}) &= E(\text{Au}_{n-1}\text{Na}) + E(\text{Au}_{n+1}\text{Na}) - 2E(\text{Au}_n\text{Na}), \\ \Delta_2 E(\text{Au}_{n+1}) &= E(\text{Au}_n) + E(\text{Au}_{n+2}) - 2E(\text{Au}_{n+1}). \end{aligned} \quad (2)$$

With the use of above formulas, we obtain $E_b = 1.27$ (AuNa), 1.38 (Au₂Na), 1.67 (Au₃Na), 1.65 (Au₄Na), 1.85 (Au₅Na), 1.80 (Au₆Na), 1.91 (Au₇Na), 1.93 (Au₈Na), and 1.98 eV (Au₉Na) for Au_nNa ($n=1-9$) clusters and 1.10 (Au₂), 1.14 (Au₃), 1.46 (Au₄), 1.60 (Au₅), 1.83 (Au₆), 1.78 (Au₇), 1.90 (Au₈), 1.89 (Au₉), and 1.97 eV (Au₁₀) for Au_{n+1} ($n=1-9$) clusters, respectively. Moreover, the trends of E_b with respect

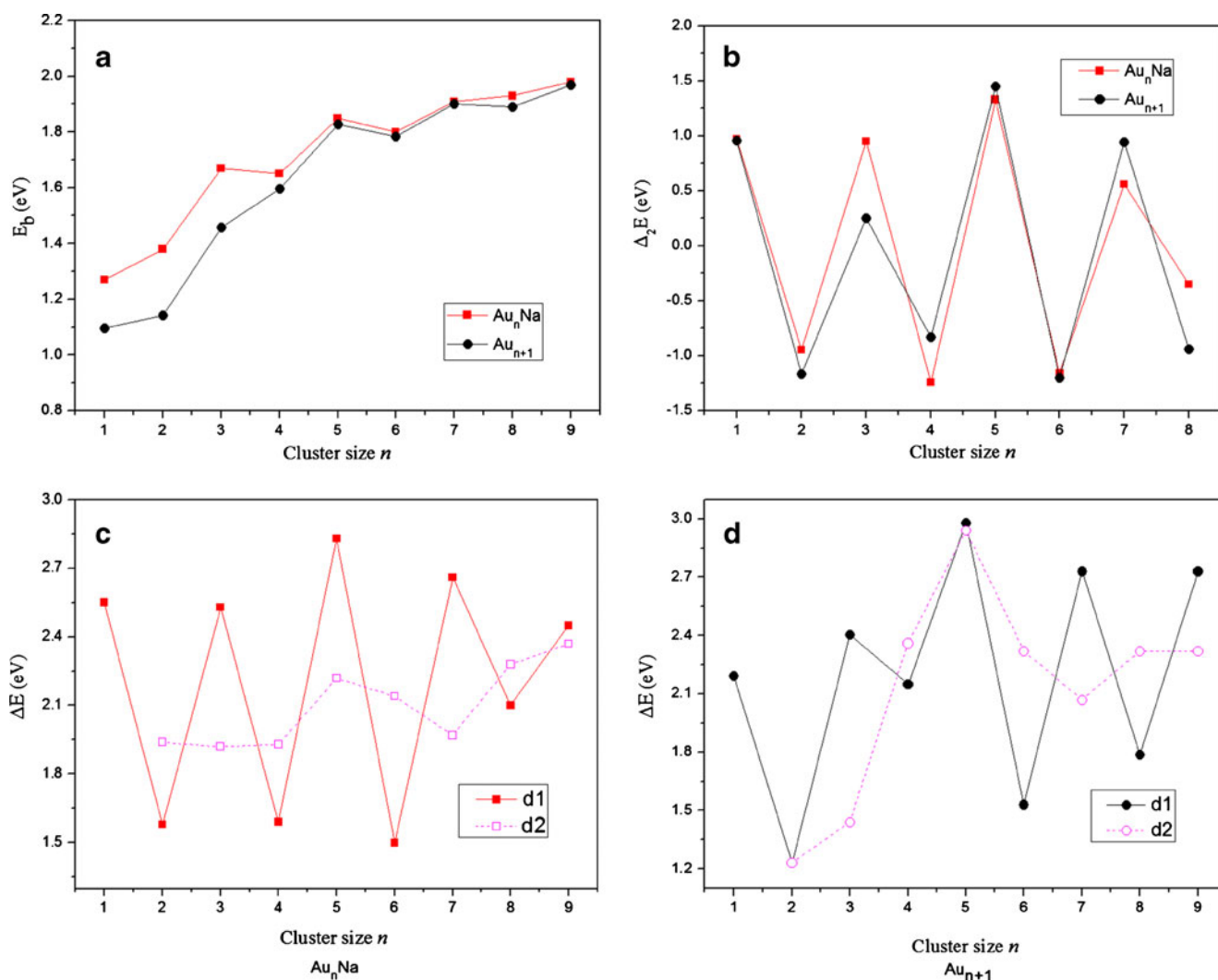


Fig. 2 Size dependence of (a) the atomic average binding energies for Au_nNa and Au_{n+1} clusters, (b) the second-order difference of energies for Au_nNa and Au_{n+1} clusters, (c) the fragmentation energies for

Au_nNa clusters, and (d) the fragmentation energies for Au_{n+1} clusters in the size range $n=1-9$

to cluster size are plotted in Fig. 2a. In Fig. 2a, the atomic average binding energies show a stair-step increasing tendency for both Au_nNa and Au_{n+1} ($n=1-9$) clusters. It illustrates that, as the cluster size grows, the addition of the gold atoms makes the number of Au-Na (Au-Au) bonds increase. Consequently, the contributions to the E_b from Au-Na (Au-Au) interaction increase, so the stabilities tend to be more stable for the large Au_nNa (Au_{n+1}) clusters. Although the E_b of Au_nNa clusters are higher than those of Au_{n+1} for $n=1-4$, the influence of impurity on the large pure clusters almost can be neglected. As presented in Fig. 2a, there are two visible peaks in the curves at $n=3$ and 5, indicating that Au_3Na , Au_5Na , and Au_6 clusters are relatively more stable than the other ones. In addition, it can be found that the E_b curves of Au_nNa and Au_{n+1} clusters increase noticeably from $n=1$ to $n=5$; then, the values show a smooth growing tendency in the cluster size range of $n=6-9$.

In view of the second-order difference of energies, positive values of Δ_2E mean that the clusters are particularly stable. Considering the dependence of the Δ_2E on the cluster size, the relationships of Δ_2E versus n are plotted in Fig. 2b. For Au_{n+1} ($n=1-9$) clusters, a pronounced odd-even oscillation can be depicted from Fig. 2b. It indicates that the Au_{n+1} clusters, at $n=3, 5$, and 7, correspond to a higher relative stability compared to the neighbors. A similar oscillatory behavior is also found for Na-doped clusters, which may be due to the fact that both Na and Au atoms possess an outermost s valence electron. Therefore, we can conclude that the clusters with even-number valence electrons are more stable than those with odd-number valence electrons. Particularly, the most stable geometries of Au_nNa and Au_{n+1} clusters are assigned to 6-electron systems (Au_5Na and Au_6) since the calculated highest values of Δ_2E (1.33 eV for Au_5Na and 1.45 eV for Au_6).

It is well known that the fragmentation energies (Δ_1E) have also proved to be a powerful tool to reflect the relative stability of cluster. The fragmentation energies with respect to single Au dissociation are $\Delta_1E(d1) = E(Au_{n-1}Na) + E(Au) - E(Au_nNa)$ for Au_nNa clusters, and $\Delta_1E(d1) = E(Au_n) + E(Au) - E(Au_{n+1})$ for Au_{n+1} clusters. In order to investigate further the dissociation channel preferred by a cluster, the cluster fragmented into the Au_2 dimer is also considered: $\Delta_1E(d2) = E(Au_{n-2}Na) + E(Au_2) - E(Au_nNa)$ and $\Delta_1E(d2) = E(Au_{n-1}) + E(Au_2) - E(Au_{n+1})$, respectively, for Au_nNa and Au_{n+1} clusters. The positive values of Δ_1E mean that the dissociation of Au atom or Au_2 dimer is an unfavorable process. Therefore, if $\Delta_1E(d1) > \Delta_1E(d2)$, the Au_2 dimer dissociation is favorable; on the contrary, $\Delta_1E(d1) < \Delta_1E(d2)$ corresponds to a preferred Au dissociation. In Fig. 2c and d, the fragmentation energies are plotted against the total number of atoms in the Au_nNa and Au_{n+1} clusters. It is easy to see that the dissociation of Au_2 dimer is favored for both series at $n=3, 5, 7, 9$, while the dissociation into a

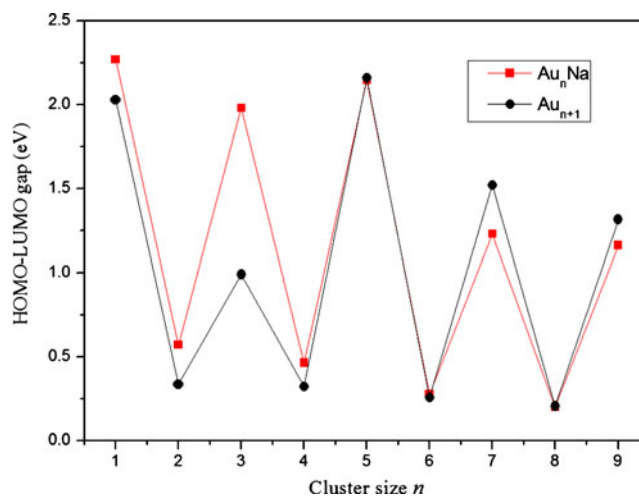


Fig. 3 Size dependence of the HOMO-LUMO energy gaps for Au_nNa and Au_{n+1} ($n=1-9$) clusters

cluster and a single Au atom is a more favorable fragmentation channel for Au_2Na , Au_4Na (Au_5), Au_6Na (Au_7), and Au_8Na (Au_9) clusters. This means that, for the clusters containing odd-number valence electrons, rather than Au_2 dimer fragmentation resulting in the formation of $Au_{n-2}Na$ (Au_{n-1}), the dissociation of Au_nNa (Au_{n+1}) into $Au_{n-1}Na$ (Au_n) is a more favorable process. Hence, it can be predicted that doped and pure gold systems have a tendency to dissociation into the smaller clusters with a closed shell, which are more stable. In addition, a distinct odd-even alternative behavior as a function of cluster size for Au_nNa and Au_{n+1} clusters is also visible in $\Delta_1E(d1)$, indicating that Au_3Na (Au_4), Au_5Na (Au_6), and Au_7Na (Au_8) clusters keep a higher stability than their vicinities, which consists with our calculated results of Δ_2E . As shown in Fig. 2c and d, interestingly, the largest Δ_1E of channels d1 and d2 are also found for the most stable configuration of Au_5Na and Au_6 ,

Table 2 Natural charge populations and natural electron configurations of Na atom in the lowest energy Au_nNa ($n=1-9$) clusters (an asterisk indicates that the value corresponds to 4s orbital)

Cluster	Natural Population	Natural electron configuration			
		2p	3s	3p	4p
$AuNa$	0.72240	6.00	0.26	0.01	0.01
Au_2Na	0.82060	6.00	0.11	0.07	0.00
Au_3Na	0.82211	6.00	0.07	0.08	0.03
Au_4Na	0.82300	6.00	0.10	0.08	0.00
Au_5Na	0.72772	6.00	0.13*	0.01	0.14
Au_6Na	0.69308	6.00	0.11	0.20	0.00
Au_7Na	0.82881	6.00	0.04	0.09	0.05
Au_8Na	0.85631	6.00	0.03	0.12	0.00
Au_9Na	0.74700	6.00	0.06	0.12	0.08

Table 3 Hardness (η), vertical ionization potential (VIP), and vertical electron affinity (VEA) of the lowest energy Au_{n+1} and Au_nNa ($n=1-9$) clusters. Experimental values of VIP for Na_{n+1} clusters are also given for comparison

Cluster size	Au_{n+1}				Au_nNa				Na_{n+1}
	η	VIP	VEA	VIP ^a	η	VIP	VEA	VIP	VIP ^d
$n=1$	3.75	9.50	2.00	9.50	3.45	7.65	0.76	4.63-8.77 ^b	4.87(02)
$n=2$	2.54	8.45	3.38	7.50	2.68	6.63	1.28		3.90(02)
$n=3$	2.72	8.28	2.76	8.60	3.53	7.69	0.64		4.24(02)
$n=4$	2.25	7.61	3.12	8.00	2.34	6.92	2.24		3.99(03)
$n=5$	3.18	8.56	2.20	8.80	3.21	8.34	1.92		4.23(03)
$n=6$	2.01	7.27	3.25	7.80	1.99	6.16	2.19	≈ 5.90 ^c	4.03(03)
$n=7$	2.59	8.04	2.87	8.65	2.43	7.00	2.14	> 6.5 ^c	4.22(03)
$n=8$	1.83	7.13	3.47	7.15	1.75	6.19	2.70	≈ 6.1 ^c	3.56(05)
$n=9$	2.33	7.60	2.94	8.20	2.29	7.20	2.63	> 6.5 ^c	3.84(03)

^a Experimental values of VIP for Au_{n+1} clusters, Ref. [46]

^b Experimental value of VIP for $AuNa$ clusters, Ref. [31]

^c Experimental values of VIP for Au_nNa clusters, Ref. [47]

^d Experimental values of VIP for Na_{n+1} clusters, Ref. [48]

so these structures are particular stable among all Au_nNa and Au_{n+1} ($n=1-9$) clusters.

The energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) is another indicator of relative stability. It reflects the ability of the molecule to participate in the chemical reaction. The cluster with a large value of HOMO-LUMO energy gap is less reactive, while the small one is related to a high chemical activity. For the low-lying configurations of Au_nNa ($n=1-9$) clusters, the HOMO and LUMO energies are summarized in Table 1; meanwhile, the HOMO-LUMO energy gaps of the ground-state isomers are plotted in Fig. 3. Figure 3 shows that the HOMO-LUMO energy gaps of Au_nNa and Au_{n+1} clusters exhibit an odd-even oscillation, and the distinct cluster intensity peaks are observed at $n=3, 5$, and 7 . Namely, the closed-shell clusters have larger HOMO-LUMO energy gaps and are relatively more chemically stable than the open-shell neighbors. In both systems, the curve shows a prominent peak for Au_5Na (Au_6), 2.15 eV (2.16 eV), manifesting they have the least chemical activities.

Population analysis

On the ground of natural population analysis (NPA) summarized in Table 2, we will describe the charge transfer of Au_nNa clusters in the following. It is found that the Na atom in the ground-state isomers carries 0.69308-0.85631 e-charges for $n=1-9$. The positive values indicate that charges in the corresponding clusters transfer from the Na atom to the Au_n frames, that is, the Na acts as electron donor in all Au_nNa clusters. This may be because the electronegativity is much larger in Au (2.54) than Na (0.93)

such that gold-sodium has a high ionic character, which coincides with the theoretical results of Tong et al. [30]. In addition, in order to understand the internal charge transfer, the natural electron configurations of impurity are also listed in Table 2. As seen in Table 2, we find that the 2p orbital behaves as a dominant core orbital in which the 6.00 electrons are full-filled. For free Na atom, the configuration of valence electron is $3s^1$. When the Na and Au atoms combine to form the Au_nNa ($n=1-9$) clusters, the NPA results illustrate that the configuration of valence electrons is $3s^{0.03-0.26}3p^{0.01-0.20}4p^{0.00-0.14}$ for sodium. Obviously, the internal electrons mainly transfer from 3s subshell to the

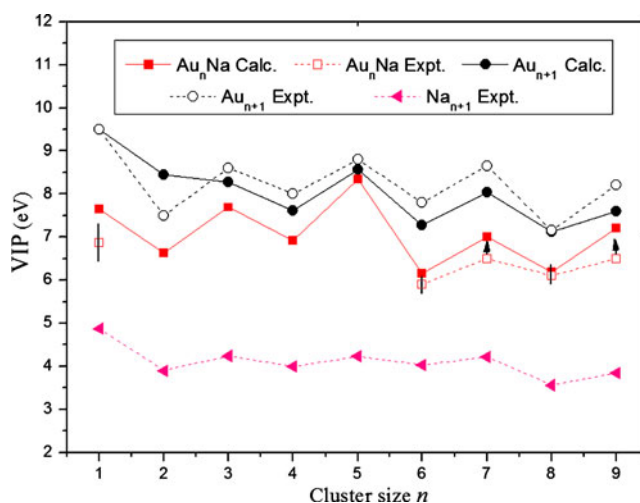


Fig. 4 Calculated and experimental vertical ionization potential (VIP) for Au_nNa and Au_{n+1} ($n=1-9$) clusters as a function of clusters size. Experimental values of VIP for Na_{n+1} clusters are also plotted. An arrow indicates that the VIP is larger or smaller than value indicated by a hollow square

vacant 3p and 4p orbitals (except for Au₅Na). Interestingly, a distinct odd-even oscillation is observed for the electrons occupying the 4p orbital, and a prominent peak corresponds to Au₅Na cluster.

Ionization potential, electron affinity, and hardness

The ionization potential ($IP = E_{\text{cation}} - E_{\text{neutral}}$) is defined as the energy difference between the ground state of the neutral and the ionized clusters. Assuming that the cation cluster has the same geometry as the optimized neutral, the IP corresponds to the vertical ionization potential (VIP). The calculated VIP results for Au_{*n*}Na and Au_{*n*+1} clusters as well as the corresponding experimental values are listed in Table 3; meanwhile, the sodium clusters are also chosen as a comparison. From Table 3, we can see that although the theoretical VIPs have a small deviation, the evolutions with the cluster size are in good agreement with the experimental results, which are presented in Fig. 4. One can find that the VIPs for each size of Au_{*n*}Na, Au_{*n*+1} and Na_{*n*+1} clusters exhibit a sequence of Au_{*n*+1} > Au_{*n*}Na > Na_{*n*+1}, so the substituted Au_{*n*} frames have enhanced the stability of sodium clusters. The odd-even alternative behaviors shown in the plot suggest that the clusters with closed electronic shell are electronically stabilized and show high VIPs. As for Au_{*n*}Na and Au_{*n*+1} clusters, meanwhile, an oscillatory phenomenon also can be observed in vertical electron affinity (VEA), which is calculated by employing the following formula: $VEA = E_{\text{optimized neutral}} - E_{\text{anion at optimized neutral geometry}}$. The clusters with smaller VEAs indicate them to be less reactive in accepting the extra electrons or more stable with respect to their neighbors.

Chemical hardness η is an indicator of the chemical reactivity, which has been established as an electronic quantity that in many cases may be used to characterize the relative stability of molecules and aggregate through the principle of maximum hardness proposed by Pearson [49]. In a finite difference approximation, we calculate the chemical hardness, that is, $\eta = (VIP - VEA)/2$. As can be seen in Table 3, the typical odd-even oscillations also can be found as the cluster size increasing, that is, the enhanced peaks are observed at the sizes $n=3, 5,$ and $7,$ which confirm the conclusions drawn from the above sections.

Conclusions

In the present DFT work, we have presented a systematic study of the geometric structures, relative stabilities, and electronic properties of Au_{*n*}Na ($n=1-9$) clusters at the PW91PW91/GEN (SDD for the Au atom and 6-311G* for

the Na atom) method level, and compared them with pure gold clusters. Our results show an early appearance of 2D→3D structural transition at the cluster of Au₆Na. In the case of Na, owing to a larger ionic radius than that of Au, the Na-peripheral and Na-capped geometries are dominant growth patterns.

The relative stabilities of Au_{*n*}Na and Au_{*n*+1} clusters are measured in terms of the atomic average binding energies, the second-order difference of energies, the fragmentation energies, and the HOMO-LUMO energy gaps, respectively. The odd-even alternations manifest that the clusters with even-number s valence electrons keep a higher stability than the clusters with the odd-number ones. For both series, it is concluded that the 6-electron systems (Au₅Na and Au₆) are the most stable configurations for $n=1-9$.

On the basis of natural population analysis, the positive charges of Na reveal that the charges transfer from Na to Au frames in gold-sodium clusters. In addition, the ionization potential, electron affinity, and chemical hardness of Au_{*n*}Na and Au_{*n*+1} clusters are discussed and compared with available experimental values. The results also confirm the stabilities of these clusters with closed electronic shell.

Acknowledgments This work was supported by the Doctoral Education Fund of Education Ministry of China (No. 20050610011) and the National Natural Science Foundation of China (No. 10974138).

References

- Kayi H (2010) J Mol Model 16:1029–1038
- Yoon B, Häkkinen H, Landman U, Wörz AS, Antonietti JM, Abbet S, Judai K, Heiz U (2005) Science 307:403–407
- Guzman J, Carretin S, Corma A (2005) J Am Chem Soc 127:3286–3287
- Tang Y, Wang SG, Li J (2010) J Mol Model. doi:10.1007/s00894-010-0793-6
- Graciani J, Oviedo J, Sanz JF (2006) J Phys Chem B 110:11600–11603
- Koszinowski K, Schröder D, Schwarz H (2003) Chem Phys Chem 4:1233–1237
- Yuan DW, Gong XG, Wu RQ (2008) Phys Rev B 78:035441
- Yuan DW, Wang Y, Zeng Z (2005) J Chem Phys 112:114310
- Cottancin E, Lermé J, Gaudry M, Pellarin M, Vialle JL, Broyer M (2000) Phys Rev B 62:5179–5185
- Zhang H, Zelmon DE, Deng L, Liu HK, Teo BK (2001) J Am Chem Soc 123:11300–11301
- Li X, Kiran B, Cui LF, Wang LS (2005) Phys Rev Lett 95:253401
- Neukermans S, Janssens E, Tanaka H, Silverans RE, Lievens P (2003) Phys Rev Lett 90:033401
- Bürgel C, Reilly NM, Johnson GE, Mitrić R, Kimble ML, Castleman AW Jr, Bonačić-Koutecký V (2008) J Am Chem Soc 130:1694–1698
- Häkkinen H (2008) Chem Soc Rev 37:1847–1859
- Häkkinen H, Moseler M, Landman U (2002) Phys Rev Lett 89:033401
- Fernández EM, Soler JM, Garzón IL, Balbás LC (2004) Phys Rev B 70:165403

17. Li XB, Wang HY, Yang XD, Zhu ZH, Tang YJ (2007) *J Chem Phys* 126:084505
18. Lee HM, Ge M, Sahu BR, Tarakeshwar P, Kim KS (2003) *J Phys Chem B* 107:9994–10005
19. Deka A, Deka RC (2008) *J Mol Struct THEOCHEM* 870:83–93
20. Knight WD, Clemenger K, Heer WAD, Saunders WA, Chou MY, Cohen ML (1984) *Phys Rev Lett* 52:2141–2143
21. Antoine R, Rayane D, Allouche AR, Frécon MA, Benichou E, Dalby FW, Dugourd P, Broyer M, Guet C (1999) *J Chem Phys* 110:5568–5577
22. Choi YC, Lee HM, Kim WY, Kwon SK, Nautiyal T, Cheng DY, Vishwanathan K, Kim KS (2007) *Phys Rev Lett* 98:076101
23. Peterson KI, Dao PD, Farley RW, Castleman AW (1984) *J Chem Phys* 80:1780–1785
24. Baletto F, Ferrando R (2005) *Rev Mod Phys* 77:371–423
25. Heer WAD (1993) *Rev Mod Phys* 65:611–676
26. Belpassi L, Tarantelli F, Sgamellotti A, Quiney HM (2006) *J Phys Chem A* 110:4543–4554
27. Joshi AM, Delgass WN, Thomson KT (2006) *J Phys Chem B* 110:23373–23387
28. Ellis JE (2006) *Inorg Chem* 45:3167–3186
29. Blaber MG, Arnold MD, Ford MJ (2010) *J Phys Condens Matter* 22:095501
30. Tong GSM, Cheung ASC (2002) *J Phys Chem A* 106:11637–11643
31. Stangassinger A, Knight AM, Duncan MA (1999) *J Phys Chem A* 103:1547–1552
32. Pelton AD (1986) *Bull Alloy Phase Diagr* 7:136–139
33. Zachwieja UZ (1993) *Z Anorg Allg Chem* 619:1095–1097
34. Watson RE, Weinert M (1994) *Phys Rev B* 49:7148–7154
35. Heiz U, Vayloyan A, Schumacher E (1996) *J Phys Chem* 100:15033–15040
36. Heinebrodt M, Malinowski N, Tast F, Branz W, Billas IML, Martin TP (1999) *J Chem Phys* 110:9915–9921
37. Heiz U, Vayloyan A, Schumacher E, Yeretizian C, Stener M, Gisdakis P, Rösch N (1996) *J Chem Phys* 105:5574–5585
38. Baruah T, Blundell SA, Zope RR (2001) *Phys Rev A* 64:043202
39. Majumder C, Kandalam AK, Jena P (2006) *Phys Rev B* 74:205437
40. Ghanty TK, Banerjee A, Chakrabarti A (2010) *J Phys Chem C* 114:20–27
41. Frisch MJ et al. (2004) Gaussian03, revision E.01. Gaussian Inc, Wallingford
42. Dolg M, Wedig U, Stoll H, Preuss H (1987) *J Chem Phys* 86:866–872
43. Schwerdtfeger P, Dolg M, Schwarz WHE, Bowmaker GA, Boyd PDW (1989) *J Chem Phys* 91:1762–1774
44. Perdew P, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Singh DJ, Fiolhais C (1992) *Phys Rev B* 46:6671–6687
45. Piacente V, Gingerich KA (1977) *High Temp Sci* 9:189
46. Jackschath C, Rabin I, Schulze W (1992) *Ber Bunsenges Phys Chem* 96:1200–1204
47. Hoshino K, Naganuma T, Watanabe K, Nakajima A, Kaya K (1993) *Chem Phys Lett* 211:571–574
48. Kappes MM, Schär M, Röthlisberger U, Yeretizian C, Schumacher E (1988) *Chem Phys Lett* 143:251–258
49. Pearson RG (1997) *Chemical hardness: applications from molecules to solids*. Wiley-VCH, Weinheim

Ab initio and DFT conformational study on *N*-nitrosodiethylamine, $(\text{C}_2\text{H}_5)_2\text{N}-\text{N}=\text{O}$

Silmar A. do Monte · Railton B. de Andrade ·
Elizete Ventura · Sidney R. de Santana

Received: 24 October 2010 / Accepted: 29 March 2011 / Published online: 28 April 2011
© Springer-Verlag 2011

Abstract *Ab initio* (MP2) and DFT (B3LYP) calculations, using the cc-pVTZ and aug-cc-pVTZ basis sets, have been performed to characterize some stationary points on the ground state potential energy surface of the title molecules. Several properties as, for instance, relative energies, the barriers for NO rotation around the NN bond, NBO charges on O and amino N atoms, as well as the dipole moments, have been calculated and analyzed in the light of the structures found. Both computational levels here employed yield three minima, in which the C_2NNO frame is ‘planar’ or ‘quasi-planar’. Important correlations between NBO charges and geometric parameters, as well as between some structural features and dipole moments are also discussed. A total of 17 structures have been found for the $(\text{C}_2\text{H}_5)_2\text{N}-\text{N}=\text{O}$ molecule. Two ranges of values have been obtained for the dipole moment, with the largest values occurring for the structures in which the nitrogen lone pair is parallel to the NO group π system. For instance, these two ranges are from ~ 4.1 to 4.5 D, and from ~ 1.6 to 2.1 D, at the MP2/cc-pVTZ level. These ranges are consistent with a larger and a smaller contribution of a dipolar resonance structure, respectively. As the method or basis set changes the values of the dipole moments change by at most ~ 0.23 D.

Keywords *Ab initio* · DFT · Nitrosamine · Structures

Introduction

The *N*-Nitrosamines ($\text{R}_2\text{N}-\text{N}=\text{O}$), are more stable than the simplest nitrosamine (H_2NNO) and are generally considered to be carcinogenic [1–5]. The first *N*-nitrosamine to be identified as carcinogenic, *N*-Nitrosodimethylamine, has been found in several common food and environmental sources (refs. [6, 7] and references therein).

The molecular structures of *N*-nitrosamines exert a pronounced influence on the bioactivity of this class of compounds [3–11]. One of the key steps to unravel how the molecular structure determines biological activity of the nitrosamines is to understand how specific substituents change their structures (see refs. [8, 9, 11] for specific examples).

The structure of the substituted *N*-Nitrosodialkylamines here studied can be represented by a resonance hybrid composed of the two structures shown in Fig. 1.

Two main important evidences for the partial double bond character of the NN bond in an *N*-Nitrosodialkylamine are the occurrence of two singlets in the ^1H NMR spectrum of *N*-Nitrosodimethylamine [12] and the planarity of their C_2NNO atoms [13, 14].

One very important structural feature of the *N*-nitrosamines is how the *R* substituents influence their properties as, for instance, the relative contributions of the two resonance hybrids. As a model for the study of the direct interaction of nitrosamines with the iron center of cytochrome P450, Richter-Addo et al. reported the synthesis and crystallographic characterization of the first nitrosamine complex of a metalloporphyrin [15–17]. These authors have observed that the coordinated *N*-Nitrosodiethylamine is best represented by a resonance hybrid having a significant contribution from the dipolar resonance hybrid (see Fig. 1). Their results also suggest that a significant contribution from the dipolar

S. A. do Monte (✉) · R. B. de Andrade · E. Ventura ·
S. R. de Santana
Departamento de Química, CCEN,
Universidade Federal da Paraíba,
João Pessoa PB 58059–900, Brazil
e-mail: silmar@quimica.ufpb.br

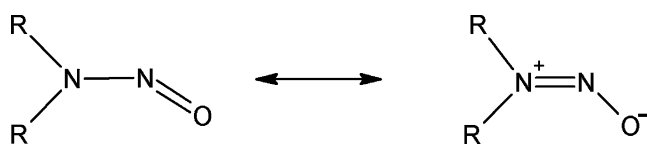


Fig. 1 Main resonance hybrids of *N*-alkyl nitrosamines

resonance hybrid (see Fig. 1) is a crucial requirement for nitrosamine coordination to Fe(III) via oxygen [18]. To the best of our knowledge there is only one paper concerning *ab initio* calculations on *N*-Nitrosodiethylamine [19]. In this work the experimental ionization potentials obtained from photoelectron spectrum are compared to the orbital energies obtained from RHF/6-31G* calculations. However, it is not clear which structures have been used for these calculations.

In this paper several stationary points on the ground state potential energy surface of the title molecule are reported. *Ab initio* (MP2) and DFT (B3LYP) methods were used. NBO charges have also been obtained and correlated to NN and NO bond lengths. The dipole moments have been qualitatively interpreted in terms of the relative contributions of the main resonance hybrids (see Fig. 1).

Computational details

All structures investigated have the same connectivity between the atoms. They have been fully optimized with or without symmetry restrictions, depending on the specific structure. After geometry optimizations frequency calculations have been performed in order to characterize the obtained stationary points. The highest possible symmetry corresponded to the C_s point group. The geometry optimizations and frequency calculations have been performed at B3LYP [20, 21] and MP2 [22, 23] levels. In the case of B3LYP calculations both cc-pVTZ and aug-cc-pVTZ [24] basis sets have been used, while at MP2 level only the former basis has been used. The use of the MP2 and B3LYP computational levels is justified if one considers the good agreement between the results obtained at these levels and the results obtained at the computationally much more demanding *ab initio* computational levels (e.g., CCSD and QCISD), in the case of $(CH_3)_2NNO$ molecule [25]. NBO charges have also been obtained, through a NBO [26] population analysis. These charges have been chosen as a more “realistic” (that is, in better agreement with the chemical sense) alternative to the Mulliken charges [27]. All calculations have been carried out using the Gaussian 03 software [28].

Results and discussions

The atom labeling scheme used for the $(C_2H_5)_2NNO$ molecule is shown in Fig. 2. The structures obtained for

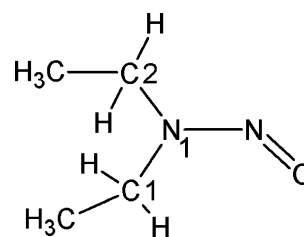


Fig. 2 Atom labeling for $(C_2H_5)_2NNO$ molecule

$(H_3C-CH_2)_2NNO$ molecule are shown in Fig. 3 (minima), 4 (‘planar’ saddle points) and 5 (‘non-planar’ saddle points). In Figs. 4 and 5 the ways the structures are connected are also shown. The normal modes of imaginary frequencies (at the MP2/cc-pVTZ level) are shown in Figs. 6 and 7, for the ‘planar’ and the ‘non-planar’ saddle points, respectively. Structures 3a, 3b and 3c¹ correspond to minima, while the remaining structures correspond to first to third order saddle points. The structure named ‘rx’ in Fig. 3 has been taken from the crystal structure of the bis(diethylnitrosamine) (*meso*-tetraphenyl-porphyrinato)iron(III) cation [15] and it is shown for comparison, since it is similar to structure 3c. Similarly to what has been found for the structures of the $(CH_3)_2NNO$ molecule [25], 3b and 3c can be obtained from 3a through ethyl rotations around the NC bonds.

In the case of structures 3d to 3k the heavy atoms are in the same plane, and they all have C_s symmetry (see Fig. 4), while the remaining structures also have C_s symmetry, but in these cases the symmetry plane only contains the NNO atoms (see Fig. 5).

The 3d structure can be obtained from 3c through rotation of both ethyl groups around the NC bonds such that all heavy atoms are in the same plane, and 3e structure differ from 3d simply by a methyl rotation around the C1C bond (see Fig. 4). It is important to mention that the value of the CC2NN dihedral angle (zero) of structure 3d is very close to the corresponding angle of the structures 3c and 3c’, which are about -1.0 and $+1.0$ degrees, respectively (see Fig. 4). The joint rotation of both ethyl groups around the NC bonds correspond to the single imaginary frequency observed for 3d, while for 3e the two imaginary frequencies correspond to rotations more localized in the ethyl group connected to the C1 atom (see Fig. 6). By following the imaginary frequency mode of the 3d structure as an intrinsic reaction coordinate (IRC) followed by a geometry optimization one can recover either the 3c structure or its *enantiomeric* pair (that is, the mirror image of the 3c structure through the plane formed by the NNO atoms, named 3c’), depending on the sense of rotation (see Fig. 4). Therefore, 3d corresponds to a transition state for a simultaneous rotation of both ethyl groups around the NC

¹ Number ‘3’ is used as a continuation of the previous nitrosamines H_2NNO (1) and $(CH_3)_2NNO$ (2), see ref. [25]

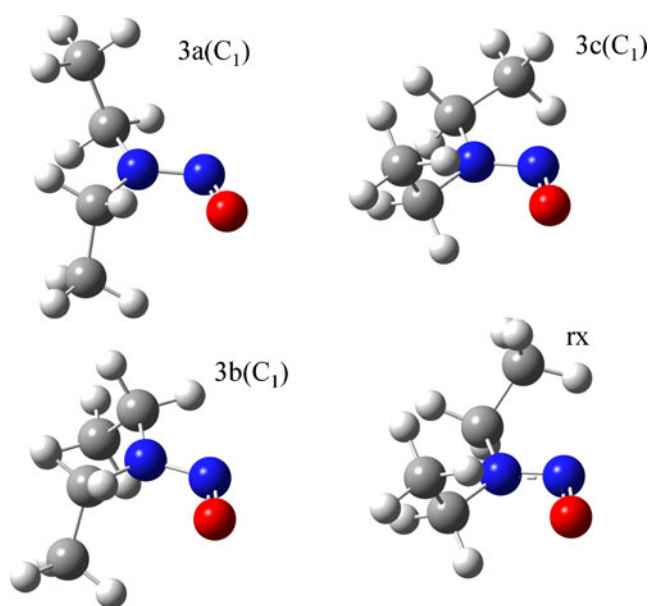


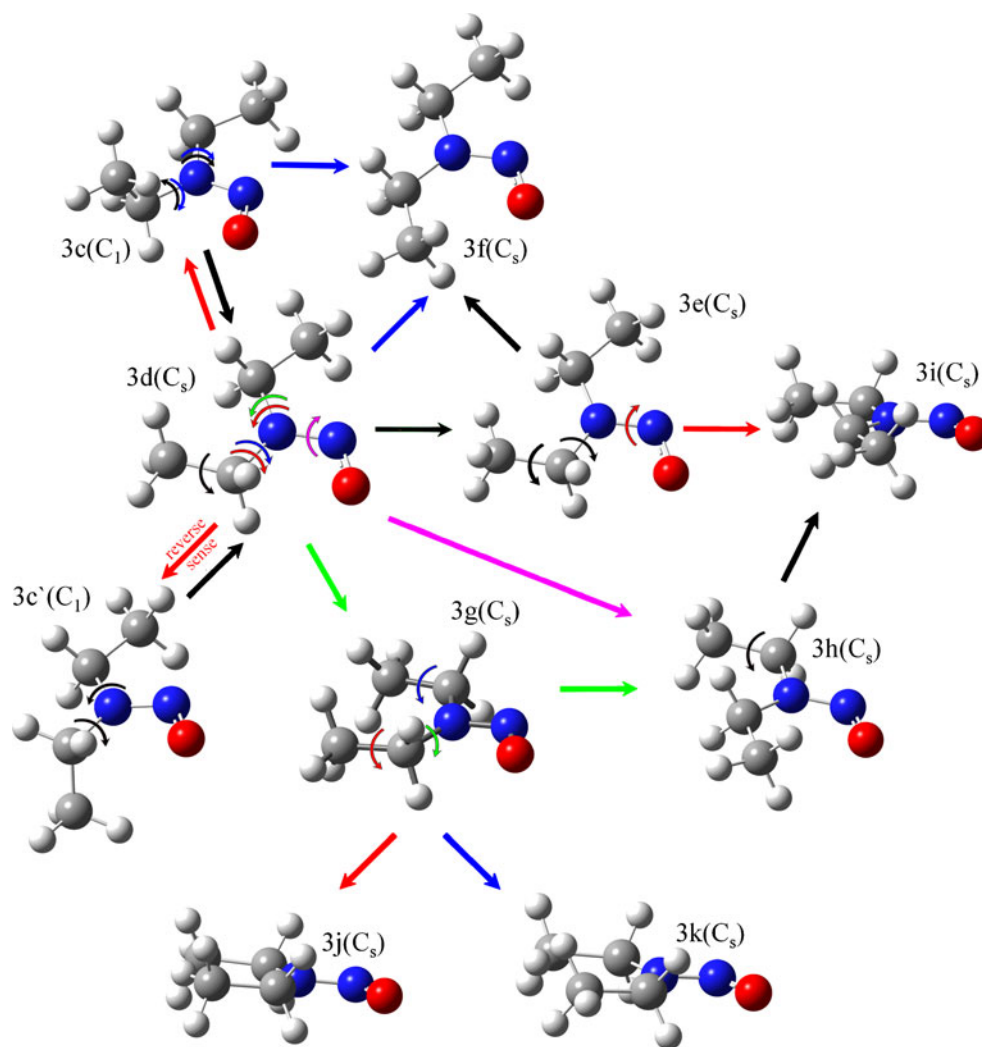
Fig. 3 Minimum energy structures obtained for the $(\text{C}_2\text{H}_5)_2\text{NNO}$ molecule, at both B3LYP and MP2 levels. The experimental structure of the $(\text{C}_2\text{H}_5)_2\text{NNO}$ ligand coordinated to the (*meso*-tetraphenyl-porphyrinato)iron(III) cation [15], obtained from x-ray, is also shown for comparison

bonds, starting from 3c or 3c' structure, that is, $3c \rightarrow [3d] \rightarrow 3c'$. On the other hand, 3f and 3g structures are derived from 3d through ethyl rotation around the NC1 and NC2 bonds, respectively (see Fig. 4). 3f can also be obtained from 3e through an ethyl rotation around the NC1 bond and a methyl rotation around the C1C bond (see Fig. 4). 3f is a first order saddle point with the imaginary frequency corresponding mainly to an ethyl rotation around the NC1 bond, while for 3g one imaginary frequency mode corresponds to an umbrella-like motion and the other corresponds to a rotation more localized around the NC1 bond (see Fig. 6). Similarly to what has been found for structure 3d, by performing an IRC scan through the imaginary frequency mode of the 3f followed by a geometry optimization one can recover either structure 3c or its *enantiomeric* pair (3c'), depending on the sense of rotation. Therefore, 3f corresponds to another transition state for ethyl rotation around the NC bonds, starting from either 3c or 3c' structure, that is, $3c \rightarrow [3f] \rightarrow 3c'$ (see Fig. 8 for the corresponding IRC plot). 3h structure may be derived either from 3g through an ethyl rotation around the NC1 bond or from 3d via NO rotation around the NN bond (see Fig. 4), and its imaginary frequencies also correspond to ethyl rotations around the NC bonds. 3i differ from 3h simply by a methyl rotation around the C2C bond (or can be obtained from 3e via NO rotation around the NN bond, as also shown in Fig. 4), while 3j and 3k differ from 3g by a methyl rotation around the C1C and C2C bonds, respectively (see Fig. 4).

In Table 1 the energetic barriers associated to conformational changes between several closely related structures (3c/3c' and 3d to 3k) are shown (see also Fig. 4). As aforementioned, two transition states, 3d and 3f, which connect 3c and its enantiomeric pair (3c') have been found. The corresponding barriers for the conformational changes $3c \rightarrow [3d] \rightarrow 3c'$ and $3c \rightarrow [3f] \rightarrow 3c'$ are 2.173 and 8.096 kcal mol⁻¹, respectively (see Table 1). Thus, the interconversion between structures 3c and 3c' via 3f involves a greater steric hindrance than the same interconversion via structure 3d. Despite the fact that structure 3e differs from 3d by just a rotation of the methyl group connected to the C1 atom (see Fig. 4) the corresponding energy difference of 3.9 kcal mol⁻¹ (see Table 1) is a bit larger than the energy difference of about 2.63 kcal mol⁻¹ between the staggered and eclipsed conformations of ethane (at the same computational level), which reflects the additional steric hindrance imposed by the CH₂ group connected to the C2 atom (see Figs. 2 and 4). The energy difference between structures 3d and 3g is mainly related to the close proximity between the two methyl groups (see Fig. 4). The smallest energy difference between these closely related structures is between structures 3g and 3h (see Table 1), which differ by a rotation of the ethyl group containing the C1 atom (see Figs. 2 and 4). The structures 3h and 3i differ by a rotation of the methyl group connected to the C2 atom, a similar situation to what has been found for structures 3d and 3e, if the NO rotation around the NN bond is taken into account. As aforementioned 3h can be generated from 3d via NO rotation, while 3i can be generated from 3e also via NO rotation. Hence the proximity between the values of the third and sixth barriers (see Table 1 and Fig. 4). The proximity between the values of the last two barriers can be explained if one considers the similarity between the structures involved, that is, both correspond to a rotational barrier of a methyl group which is close to another methyl group. Of note, of all eight 'planar' saddle point structures (that is, 3d to 3k), 3g, 3j and 3k have the two methyl groups nearest to each other. Hence the proximity between the values of the last two barriers (see Table 1).

The remaining structures (3l to 3q) also have C_s symmetry, but in these cases the symmetry plane coincides with the plane formed by the NNO atoms (see Fig. 5). All these structures have at least one imaginary frequency corresponding mainly to a NO rotation around the NN bond (see Fig. 7). Structures 3l and 3n are connected to 3f in a simple way. If, starting from 3f, the NO group is rotated around the NN bond (which causes the appearance of this rotation as an imaginary frequency) the conjugation between the nitrogen lone pair and the π system of the NO group is lost, leading to a pyramidalization around the N1 atom (that is, the two CN bonds as well as the NN

Fig. 4 ‘Planar’ saddle-point structures (3d to 3k) obtained for the $(C_2H_5)_2NNO$ molecule, at both B3LYP and MP2 levels. The relations between the obtained structures are also shown. In the case of structure 3d the rotation shown practically corresponds to the reaction coordinate (see text for details) and leads to the structure 3c, while the inverse rotation lead to its enantiomeric pair, 3c’



single bond are directed along a pyramid vertices). Such pyramidalization destroys the imaginary frequency of 3f (which, as aforementioned, corresponds mainly to a rotation around the NC1 bond) and yields two possibilities concerning the relative position of the nitrogen lone pair. In one situation (that is, of structure 3l) the lone pair is on the same side of the oxygen atom, while in the other situation (structure 3n) it is on the opposite side of the oxygen atom (see Fig. 5). 3l and 3n have the NO rotation around the NN bond as the unique imaginary frequency mode and all heavy atoms (excepting the NO group, see Fig. 5) almost share the same plane. Similarly to what has been found previously, by performing an IRC scan through this imaginary frequency mode (for both 3l and 3n structures) one can recover either the 3c structure or its *enantiomeric* pair (3c’), depending on the sense of rotation (see Fig. 8 for both IRC plots). As in the case of 3l and 3n structures, 3m and 3o can be obtained from 3g (which is a second order saddle point) through NO rotation around the

NN bond and also represent different relative positions between the N1 lone pair and the oxygen atom (see Fig. 5). The two normal modes of imaginary frequency of 3g structure can be described mainly as an ethyl rotation around the NC1 bond and an ethyl rotation around the NC2 bond (see Fig. 6). On going from 3g to 3m or 3o structures these two normal modes become absent and are replaced by two other. One of them is, as aforementioned, a NO rotation around the NN bond, and the other is a symmetry-breaking mode, located on the ethyl groups (see Fig. 7). Finally, the 3p and 3q structures can be obtained from 3b through a NO rotation around the NN bond along with a symmetry plane (containing the NNO atoms) imposition. For the former structure the nitrogen lone pair is closer to the oxygen than in the case of the latter, as can be seen from Fig. 5. As can be seen from Fig. 7, 3p and 3q have, apart from the NO rotation, an additional symmetry-breaking vibrational mode (located on the ethyl groups) of imaginary frequency. 3q is the only structure whose number of imaginary frequency

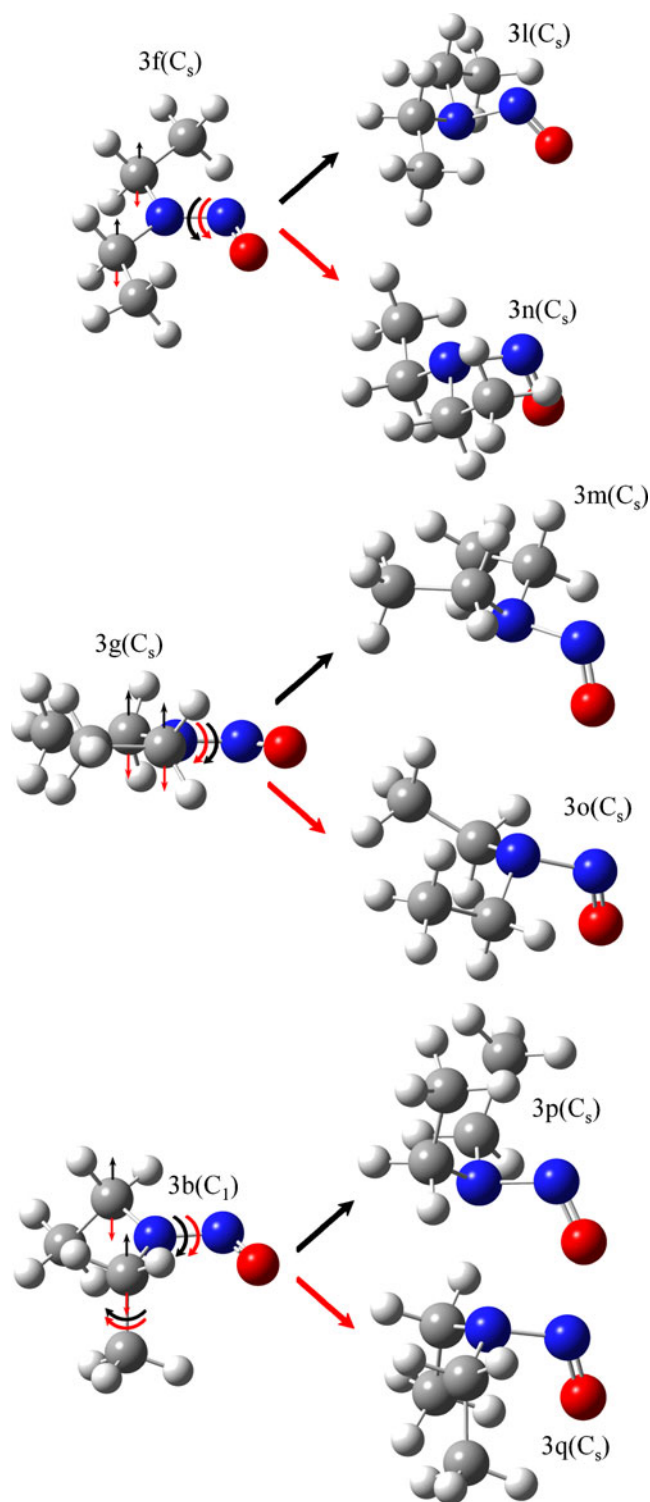


Fig. 5 ‘Non-planar’ saddle-point structures (3l to 3q) obtained for the $(\text{C}_2\text{H}_5)_2\text{NNO}$ molecule, at both B3LYP and MP2 levels. The relations between the obtained structures are also shown. The vertical arrows indicate the pyramidalization around the N1 atom as a result of geometry optimization, while the methyl rotation in structure 3b is such that a symmetry plane containing the NNO atoms (upon the transformations $3b \rightarrow 3p$ and $3b \rightarrow 3q$) is imposed

vibrational modes depends on the method, that is, at the B3LYP/cc-pVTZ level the aforementioned symmetry-breaking vibrational mode is absent as a mode of imaginary frequency. Since structures 3c (or 3c') can be obtained from the 3l or 3n structures through an IRC scan mainly composed of the NO rotation around the NN bond (see Fig. 8), followed by geometry optimization, one can conclude that, of all ‘non-planar’ structures found (3l to 3q), only 3l and 3n can be considered true transition states (at the MP2/cc-pVTZ level) for the rotational barrier of the NO group around the NN bond, in the case of ethyl substituents.

In Table 2 the energetic barriers associated to conformational changes between several closely related structures (3b, 3f, 3g and 3l to 3q) are shown (see also Figs. 4 and 5). All energetic barriers shown in this table are larger than the barriers shown in Table 1, a result that can be attributed to the lost of planarity of the C_2NNO frame as the rotations specified in Fig. 5 (of structures 3b, 3f and 3g), followed by geometry optimization, are done. The energy differences between structures 3c and 3l (of $19.386 \text{ kcal mol}^{-1}$) and between structures 3c and 3n (of $22.782 \text{ kcal mol}^{-1}$) correspond to the barriers for the following conformational changes, $3c \rightarrow [3l] \rightarrow 3c'$ and $3c \rightarrow [3n] \rightarrow 3c'$ (see Fig. 8 for the corresponding IRC plots), that is, for interconversion between structures 3c and 3c' through NO rotation around the NN bond. As expected, the last two energetic barriers shown in Table 2 have the largest values, since the NO rotation done in structure 3b (followed by geometry optimization with a symmetry plane imposition, in order to yield structures 3p or 3q) increases considerably the steric hindrance between the two methyl groups, as a result of the pyramidalization around the N1 atom. Such pyramidalization is accompanied by a decrease of the $\langle \text{CNC} \rangle$ angle, approximating the two methyl groups even more (Fig. 5 shows how the structure 3b is related to structures 3p and 3q).

The geometrical parameters of $(\text{C}_2\text{H}_5)_2\text{NNO}$ structures, obtained from MP2/cc-pVTZ calculations, are shown in Table 3. The calculated values for the NN, NO and NC1 bond lengths show small variations, for structures 3a to 3k (which are the structures where the heavy atoms share (or almost share) the same plane). In the case of the NN bond length the results obtained range from: 1.326 \AA (3d structure) to 1.333 \AA (3j and 3k), while for the NO bond the calculated lengths range from 1.230 \AA (3h) to 1.239 \AA (3d and 3e). These ranges are, as expected, very close to those obtained for structures 2a to 2d (which are the structures where the heavy atoms are in the same plane) of $(\text{CH}_3)_2\text{NNO}$, that is, 1.323 to 1.335 \AA and 1.229 to 1.236 \AA , respectively [25].

Fig. 6 Values of imaginary frequency modes (in cm^{-1}) obtained for the 'planar' structures 3d to 3k, at the MP2/cc-pVTZ level

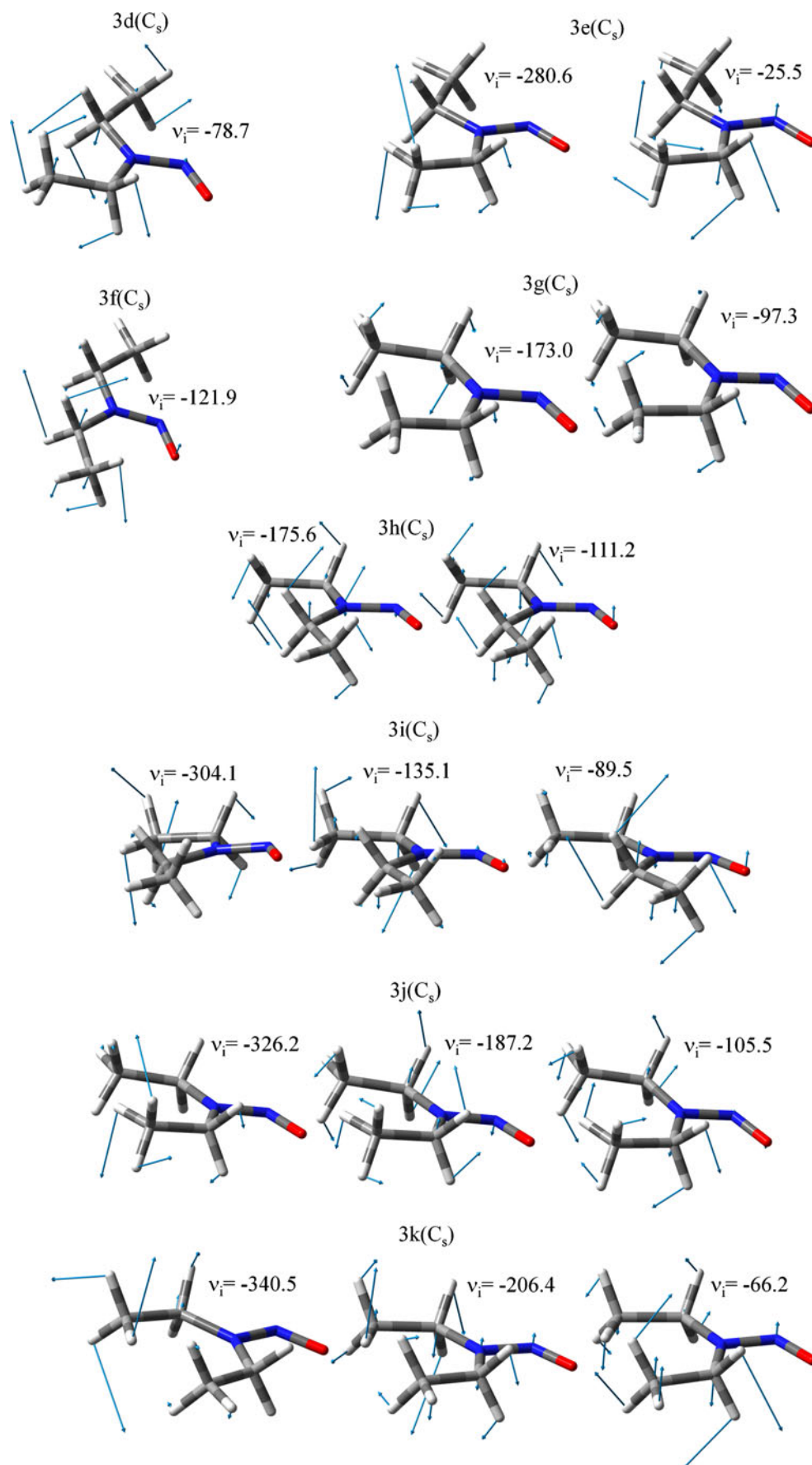
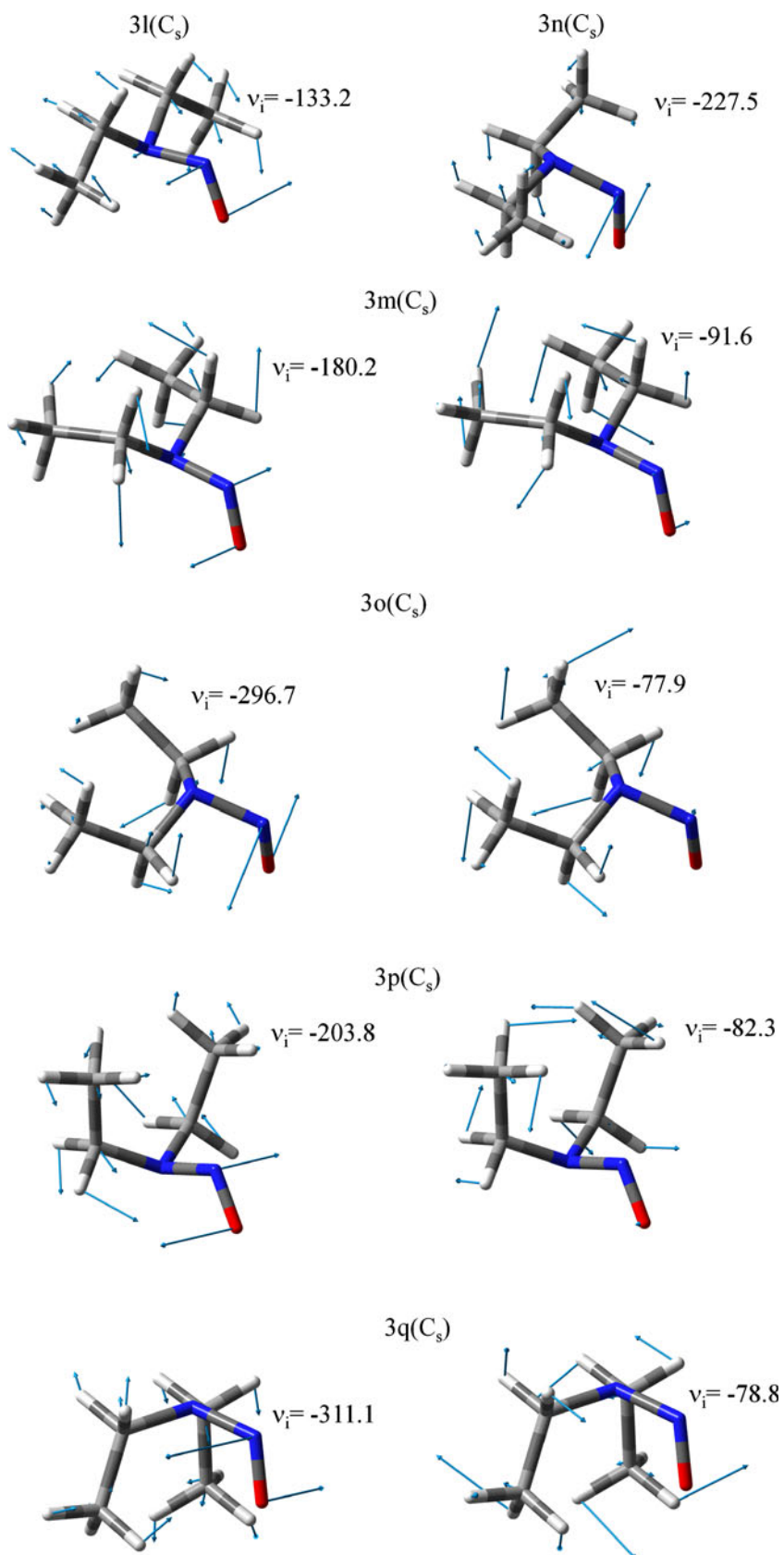


Fig. 7 Values of imaginary frequency modes (in cm^{-1}) obtained for the ‘non-planar’ structures 3l to 3q, at the MP2/cc-pVTZ level



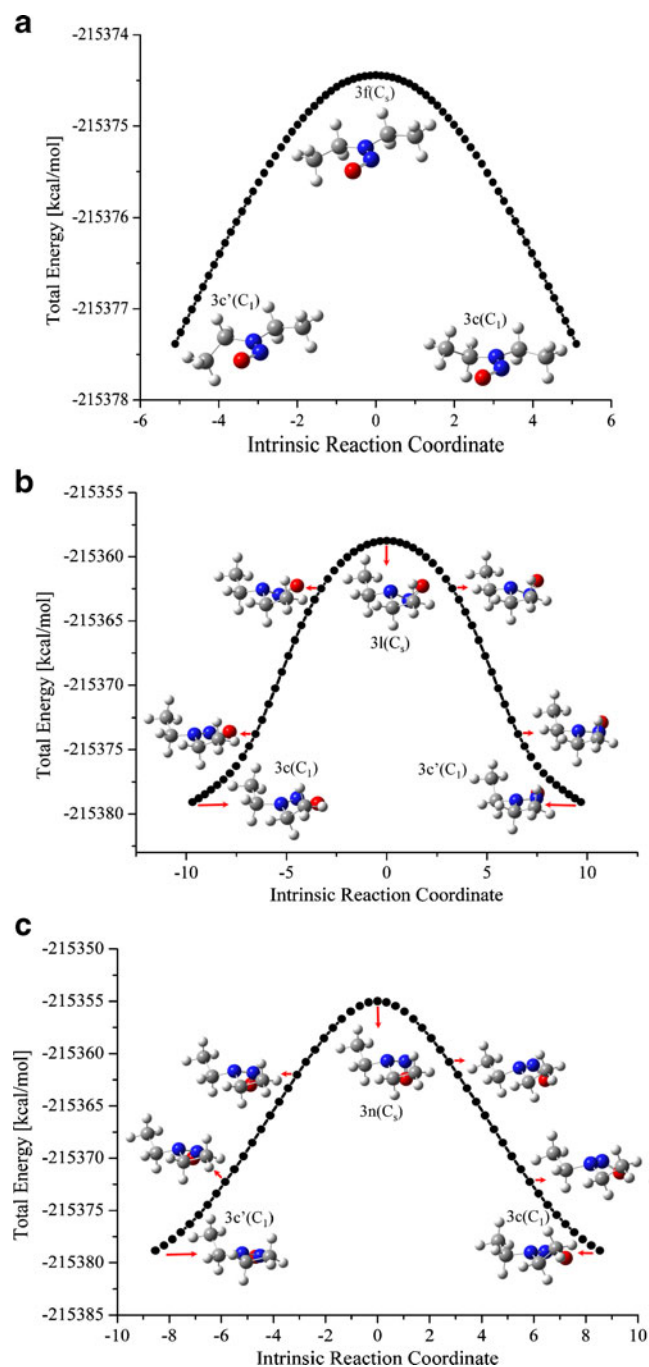


Fig. 8 IRC plot showing the connection between structures 3f \rightarrow 3c/3c' (a), 3l \rightarrow 3c/3c' (b) and 3n \rightarrow 3c/3c' (c). For structure 3f the reaction coordinate (see Figs. 4 and 6) is mainly composed of a rotation of the ethyl group closer to the oxygen atom. For the imaginary frequency modes of the structures 3f, 3l and 3n refer to Figs. 6 and 7

The NC1 bond lengths obtained for structures 3a to 3k are significantly higher than the bond lengths obtained for the corresponding structures of the $(\text{CH}_3)_2\text{NNO}$ molecule (that is, 2a to 2d), probably due to the steric hindrance imposed by the additional methyl groups. While for the

structures of the former molecule the values of the NC1 bond length range from 1.451 (3a) to 1.471 Å (3h), for the latter molecule they range from 1.444 to 1.448 Å [25].

In the case of the NC2 bond lengths the variation observed for structures 3a to 3k is even larger, varying from 1.448 (3b) to 1.483 Å (3k). This set of structures can be subdivided into two subsets, according to the values obtained for the NC2 bond length. The first set includes structures 3a to 3f, with bond lengths ranging from 1.448 (3b) to 1.458 Å (3f), while the other set includes the remaining structures, which have bond lengths varying from 1.475 (3h) to 1.483 Å (3k). This second set is such that the ethyl rotation around the NC2 bond positions the methyl group connected to the C2 atom farther from the NO group than in the case of the previous 3d, 3e and 3f structures (see Fig. 4).

If one compares the values of the $\langle \text{C}_{\text{NN}} \rangle$, $\langle \text{C}_{1\text{NN}} \rangle$ and $\langle \text{C}_{2\text{NN}} \rangle$ bond angles obtained for the actual structures with the values obtained for the structures of the $(\text{CH}_3)_2\text{NNO}$ molecule [25] one can estimate the steric effect imposed by the additional methyl groups. In the case of the 'planar' structures of $(\text{CH}_3)_2\text{NNO}$ these angles vary from 120.8 to 124.6 degrees ($\langle \text{C}_{\text{NC}} \rangle$), 119.6 to 123.2 degrees ($\langle \text{C}_{1\text{NN}} \rangle$) and 115.1 to 117.0 degrees ($\langle \text{C}_{2\text{NN}} \rangle$), at the MP2/cc-pVTZ level. The corresponding angles of structures 3a to 3e are approximately in the same range, while structures 3f to 3k show greater variations in the values of these angles (see Table 3 and Fig. 4). For instance, in the case of 3f structure the proximity between the methyl and NO groups leads to an increase in the $\langle \text{C}_{1\text{NN}} \rangle$ angle of at least 5 degrees, while the $\langle \text{C}_{\text{NC}} \rangle$ angle decreases by about the same quantity, as compared to the corresponding values of the 'planar' structures of $(\text{CH}_3)_2\text{NNO}$ [25]. The proximity between the methyl groups of structures 3g, 3j and 3k (see Fig. 4) is responsible for a substantial increase of almost 10 degrees in the $\langle \text{C}_{\text{NC}} \rangle$ angle, along with a decrease of at least 3.6 and 5 degrees, respectively, in the values of the $\langle \text{C}_{1\text{NN}} \rangle$ and $\langle \text{C}_{2\text{NN}} \rangle$ angles. With the exception of structures 3f, 3h and 3i, the values obtained for the $\langle \text{NNO} \rangle$ angle of structures 3a to 3k are approximately in the same range obtained for structures 2a to 2d [25], that is, from ~ 114.0 to 116.0 degrees. In the case of 3f, 3h and 3i structures the methyl group connected to the C1 atom (see Fig. 4) is closer to the oxygen atom, leading to slightly larger values for the $\langle \text{NNO} \rangle$ angle, of ~ 118.0 degrees.

The remaining structures, that is, 3l to 3q, have the N1 lone pair perpendicular to the NO group π system, leading, as aforementioned, to greater NN bond lengths as well as to smaller NO bond lengths. Such lack of conjugation leads, as expected, to a pyramidalization of the groups around the N1 atom (the nitrogen atom bonded to the C atoms, see Fig. 2). Therefore, the $\langle \text{C}_{\text{NC}} \rangle$, $\langle \text{C}_{1\text{NN}} \rangle$ and $\langle \text{C}_{2\text{NN}} \rangle$ angles of these structures tend to be smaller than the angles of the

Table 1 Energetic barriers, in kcal mol⁻¹ (obtained at MP2/cc-pVTZ level) associated to conformational changes between several closely related structures. These barriers refer to the ‘planar’ structures

themselves (3d to 3k) and to the ‘planar’ structures and the 3c (which is quasi-planar) structure of (H₃C-CH₂)₂NNO (see also Fig. 4)

Related structures							
3c→3d*	3c→3f*	3d→3e	3d→3 g	3 g→3 h	3 h→3i	3 g→3j	3 g→3k
2.173	8.096	3.904	7.816	1.727	4.309	5.655	5.975

*Genuine transition states

previous structures, with main exceptions observed for structures 3p and 3q, due to the steric hindrance imposed by the methyl groups. The values of the \angle_{NNO} angles calculated for the structures 3l to 3q also tend to be smaller than the values obtained for the previous structures (see Table 3), similarly to what has been found for both structures corresponding to transition states for NO rotation around the NN bond of the (CH₃)₂NNO molecule [25].

The experimental geometrical parameters of the (C₂H₅)₂NNO ligand coordinated to the (*meso*-tetraphenylporphyrinato)iron(III) cation [15] is shown in Table 3 and its structure is shown in Fig. 3 for comparison. As this figure shows the structure of the coordinated *N,N*-diethylnitrosamine is similar to the 3c structure. Although the geometrical parameters here obtained cannot be directly compared to the geometrical parameters of the coordinated ligand, since the coordination as well as the crystalline state are expected to alter the structure of this ligand, the modifications observed for the NO and NN bond lengths, as well as the \angle_{NNO} bond angle, are expected to be mainly due to the coordination. In fact, these parameters are the ones which show the greatest variations, as one compares the coordinated structure to the gas-phase free ligand (see Table 3). Thus, the NN and NO bond lengths decreases and increases, respectively, by 0.05 Å and 0.022 Å, while the NC1 and NC2 bond lengths decreases and increases, respectively, by 0.012 Å and 0.014 Å. The \angle_{NNO} bond angle decreases by ~5 degrees, while the other bond angles vary by at most 1.8 degrees. From Fig. 3 one can also clearly see that the \angle_{CC2NN} dihedral angle increases upon coordination, probably due to packing forces in the crystal. The actual difference, obtained from comparison between the dihedral angle of the MP2/aug-cc-pVTZ optimized structure and the coordinated ligand is ~33 degrees.

The relative energies of structures 3a to 3q, calculated at B3LYP level with the cc-pVTZ and cc-pVTZ basis sets and at MP2 level with the cc-pVTZ basis set, are shown in Fig. 9. While the greatest energy difference observed for the (CH₃)₂NNO molecule, at the MP2 level, is ~23.9 kcal mol⁻¹ (between structures 2a and 2f) [25], in the case of (C₂H₅)₂NNO molecule it is ~31.7 kcal mol⁻¹ (between structures 3a and 3q), due to the steric hindrance imposed by the additional methyl groups (see Fig. 5). As can be seen from Fig. 9 there is a gradual increase in the energy of the structures, at the B3LYP level. However, some structures which are very close in energy (that is, (3i, 3j, 3k) and (3m, 3n), see Figs. 4 and 5), at the B3LYP level, have their energy ordering reversed, at the MP2 level. It is worthy to mention that for most structures which have the C2NNO atoms sharing the same plane (that is, 3a to 3k) the relative energies at the MP2/cc-pVTZ level are higher than the corresponding energies at the B3LYP/cc-pVTZ level. On the other hand, for the remaining structures the reverse is true. The basis set effect, at the B3LYP level, is very small.

The energies of the obtained three minima (3a, 3b and 3c, see Figs. 3 and 9) are very close, differing by at most ~1.5 kcal mol⁻¹. The energy difference between structures 3l/3n and 3c (of ~19.4 kcal mol⁻¹ and ~22.8 kcal mol⁻¹, respectively) correspond to true rotational barriers for NO rotation around the NN bond for the (C₂H₅)₂NNO molecule (see Table 1). On the other hand, the energy difference between structures 3b and 3q (of ~31.1 kcal mol⁻¹, at MP2/cc-pVTZ level (see Table 1) and ~32.5 kcal mol⁻¹, at B3LYP/cc-pVTZ) corresponds to a true rotational barrier for NO rotation around the NN bond only at the B3LYP level, since only at this level this rotation (followed by geometry optimization) yields structure 3b and corresponds to the unique imaginary frequency.

Table 2 Energetic barriers, in kcal mol⁻¹ (obtained at MP2/cc-pVTZ level) associated to conformational changes between several closely related structures of (H₃C-CH₂)₂NNO (see also Fig. 5)

Related structures							
3f*→3l*	3f*→3n*	3 g→3 m	3c→3l*	3c→3n*	3 g→3o	3b→3p	3b→3q
11.290	14.686	13.483	19.386	22.782	15.54	28.765	31.605

*Genuine transition states

Table 3 Geometrical parameters obtained from MP2/cc-pVTZ calculations, for structures of $(C_2H_5)_2NNO$. Angles are in degrees and lengths in Å. In the case of dihedral angles the numbers shown correspond to the absolute values

	3a	3b	3c	rx ^a	3d	3e	3f	3g	3h
NN	1.329	1.330	1.326	1.276	1.326	1.327	1.330	1.332	1.331
NO	1.235	1.235	1.238	1.260	1.239	1.239	1.231	1.237	1.230
NC1	1.451	1.452	1.452	1.440	1.458	1.461	1.468	1.463	1.471
NC2	1.449	1.448	1.454	1.468	1.456	1.456	1.458	1.476	1.475
<CNC	122.4	122.6	120.1	120.1	123.7	124.7	114.1	133.9	120.4
<C1NN	121.3	121.0	121.0	122.8	118.5	117.9	129.9	116.0	129.3
<C2NN	116.2	116.3	118.8	117.0	117.8	117.3	116.1	110.1	110.4
<NNO	114.7	114.5	114.4	109.5	113.9	114.1	117.2	115.1	118.0
<C1NNO	0.5	2.3	0.2	1.1	0.0	0.0	0.0	0.0	0.0
<CC1NN	81.4	79.7	81.9	81.8	180.0	180.0	0.0	180.0	0.0
<CC2NN	106.0	117.3	1.0	32.7	0.0	0.0	0.0	180.0	180.0
	3i	3j	3k	3l	3m	3n	3o	3p	3q
NN	1.331	1.333	1.333	1.494	1.505	1.522	1.507	1.480	1.462
NO	1.231	1.237	1.237	1.194	1.191	1.196	1.197	1.198	1.207
NC1	1.468	1.467	1.465	1.472	1.472	1.468	1.470	1.469	111.6
NC2	1.481	1.479	1.483	1.472	1.472	1.468	1.470	1.469	111.6
<CNC	121.3	135.6	135.6	109.4	113.0	110.4	114.7	116.8	119.9
<C1NN	129.0	115.3	115.2	103.1	102.9	107.0	105.7	107.2	111.6
<C2NN	109.7	109.1	109.1	103.1	102.9	107.0	105.7	107.2	111.6
<NNO	118.1	115.1	115.1	112.1	111.2	112.7	113.3	110.8	113.9
<C1NNO	0.0	0.0	0.0	123.1	121.1	59.2	61.0	116.9	68.6
<CC1NN	0.0	180.0	180.0	74.4	160.7	66.5	159.0	53.1	66.3
<CC2NN	180.0	180.0	180.0	74.4	160.7	66.5	159.0	53.1	66.3

^a Geometrical parameters taken from the x-ray structure of the ligand coordinated to the (*meso*-tetraphenyl-porphyrinato) iron(III) cation [15]

The calculated dipole moments are shown in Fig. 10. As in the case of the *N*-Nitrosodimethylamine [25], two ranges of values have been obtained for the dipole moment of $(C_2H_5)_2NNO$ molecule, one from ~ 3.9 to 4.5 D and the other from ~ 1.6 to 1.9 D (at the MP2/cc-pVTZ level). The first range comprises the structures from 3a to 3k, in which

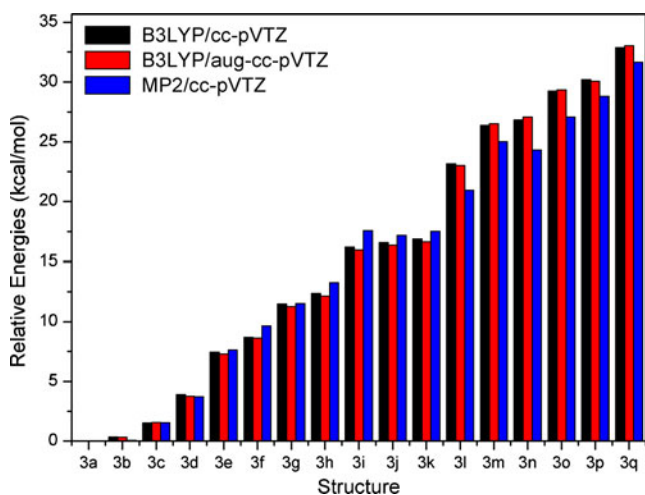


Fig. 9 Relative energies for the structures of the $(C_2H_5)_2NNO$ molecule, obtained at the B3LYP/cc-pVTZ, B3LYP/aug-cc-pVTZ and MP2/cc-pVTZ levels. The ZPE values have been computed with the cc-pVTZ basis set using the vibrational frequency scaling factor of 0.965 and 0.950 for B3LYP and MP2 methods, respectively [30]

the N1 lone pair is parallel (or approximately so) to the π system of the NO group, while the second range includes the remaining structures (3l to 3q), which have the N1 lone pair perpendicular to the π system of the NO group. The basis set effect, in the case of calculated dipole moments, is more pronounced than in the case of energy differences, and leads to an increase of ~ 0.2 D as the basis set is increased from cc-pVTZ to aug-cc-pVTZ. The differences between B3LYP and MP2 results are almost negligible.

As in the case of the $(CH_3)_2NNO$ molecule [25], a linear correlation between the calculated NN and NO bond lengths and the NBO charges on N1 and O atoms has been observed, as shown in Fig. 11a and b below (at the MP2/cc-pVTZ level). The values of R^2 for the linear correlation between the NN/NO bond lengths and the NBO charge on N1/O are 0.978 and 0.977, respectively. Despite the fact the R^2 values are very similar, the standard errors for the intercepts and the slopes are, respectively, equal to 9.58×10^{-3} Å and 3.298×10^{-2} Å/charge (for the NN \times NBO charge on N1 plot) and 2.72×10^{-3} Å and 8.09×10^{-3} Å/charge (for the NO \times NBO charge on O plot). These results suggest that, in the case of the *N*-Nitrosodiethylamines, that the NBO charge on O is a more suitable candidate for a structural probe, a result also obtained for the *N*-Nitrosodimethyl amines [25]. As in the case of $(CH_3)_2NNO$ molecule [25] the values of the NBO charges obtained from

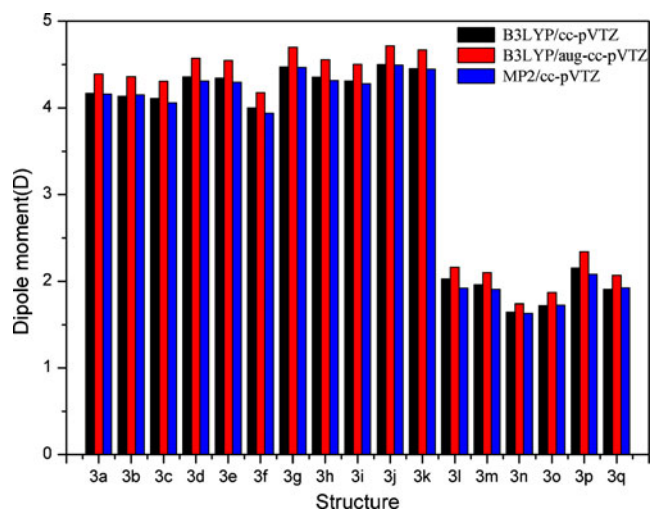


Fig. 10 Calculated dipole moments for structures 3a to 3q of $(C_2H_5)_2NNO$ molecule

these correlations are consistent with the relative contributions of the resonance structures shown in Fig. 1, that is, shorter NN (or NO) bond lengths imply in less negative charges on N1 (or O); conversely, the longer the NN (or NO) bond lengths, more negative the charges on N1 (or O, see Fig. 11a and b). A similar type of correlation has been obtained for *p*-substituted derivatives of nitrosobenzene and *p*-substituted *N,N*-dimethylanilines [29].

Conclusions

Several structures of the molecule *N*-Nitrosodiethylamine have been studied at *ab initio* (MP2) and DFT (B3LYP) levels. The structures studied for each molecule are such that the connectivity between the atoms has been maintained.

In total seventeen structures have been obtained, corresponding to three minima, four first-order saddle points, seven second-order saddle points and three third-order saddle points, at the MP2/cc-pVTZ level. For the three minima, the C1, C2, N1, and NO atoms almost share the same plane (see Figs. 2 and 3). Two of these three minima correspond to different relative locations of the terminal methyl groups. In one case they are on the same side of the aforementioned plane, while for the other minimum they are on opposite sides (see Fig. 3). The third minimum is closely related to the structure of the *N*-nitrosodiethylamine ligand coordinated to the (*meso*-tetraphenyl-porphyrinato) iron(III) cation [15]. One very important finding is that no minimum energy structure in which the C_2NNO frame deviates significantly from planarity has been obtained, which suggests that such planarity (or nearly planarity) between these atoms is an important stabilizing factor for the *N*-dialkylnitrosamines.

Eight of the 17 structures (3d to 3k, see Fig. 4) have C_s symmetry and also have the heavy atoms sharing the same plane (the symmetry plane), and are interconnected through rotations around the NC or C1C (that is, ethyl or methyl rotations) bonds. Two of these eight structures are first order saddle points, while the number of second and third order saddle points is the same, that is, three. These two first-order saddle points (structures 3d and 3f, see Fig. 4) corresponds to transition states for ethyl rotations around the NC bonds, starting either from the 3c structure (see Fig. 4) or from its enantiomeric pair (named 3c'). The remaining six structures also have C_s symmetry, but now the symmetry plane only contains the NNO atoms (see Fig. 5). They also have as the common characteristic the orthogonality between the nitrogen lone pair and the π system of the NO group. Two of these six structures correspond to first-order saddle points, while the remain-

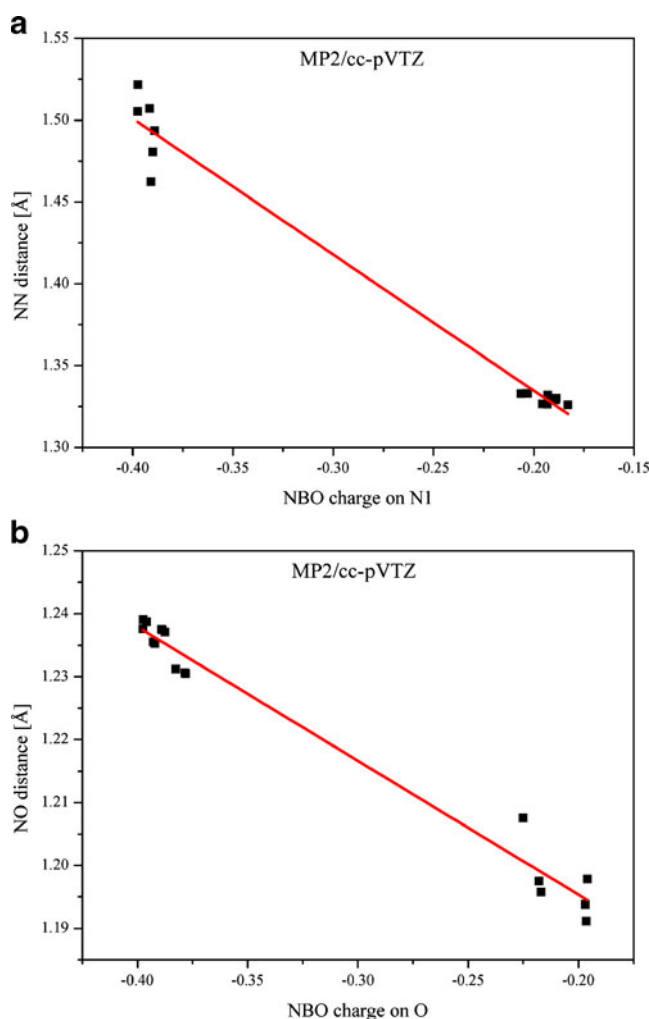


Fig. 11 Linear correlation plot of (a) NN length \times NBO charge on N1 (b) NO length \times NBO charge on O for $(C_2H_5)_2NNO$ molecule, calculated at the MP2/cc-pVTZ level

ing four correspond to second-order saddle points. All six structures have at least one imaginary frequency vibrational mode corresponding mainly to a NO rotation around the NN bond. Two structures (that is 3l and 3n, see Fig. 5) are true transition states for the interconversion between structures 3c and 3c' (its enantiomeric pair) via NO rotation around the NN bond, at both MP2/cc-pVTZ and B3LYP/cc-pVTZ levels, and the corresponding barriers are ~ 19.4 and 22.8 kcal mol $^{-1}$. The corresponding values for the (CH $_3$) $_2$ NNO molecule are ~ 21.5 and 23.9 kcal mol $^{-1}$, at the MP2/cc-pVTZ level [25]. On the other hand, only at the B3LYP/cc-pVTZ level the 3q structure is also a transition state for NO rotation around the NN bond.

A very interesting linear correlation between bond lengths and NBO charges of the 17 structures of the (C $_2$ H $_5$) $_2$ NNO molecule has been observed, that is, NN bond length versus NBO charge on N1 atom and NO bond length versus NBO charge on oxygen atom. Although a better linear correlation is obtained in the latter case, both results open the possibility of using NBO charges as structural probes, at least in the case of *N*-Nitrosodialkylamines. Such linear correlation is also consistent with the expected relative contribution of the resonance hybrids (see Fig. 1). This type of linear correlation has also been obtained for the (CH $_3$) $_2$ NNO molecule [25] and a similar type of correlation has been obtained for *p*-substituted derivatives of nitrosobenzene and *p*-substituted *N*, *N*-dimethylanilines [29].

Two ranges of values have been obtained for the dipole moment. The greatest values occur for the structures in which the nitrogen lone pair is parallel to the NO group π system (that is, for the 'planar' structures, see Fig. 4), and are from ~ 3.9 to 4.5 D, and from ~ 1.6 to 1.9 D (for the 'non-planar' structures, see Fig. 5), at the MP2/cc-pVTZ level. The corresponding ranges in the case of the (CH $_3$) $_2$ NNO molecule are very similar [25], varying from ~ 4.0 to 4.3 D, and from ~ 1.7 to 2.0 D, at the same computational level.

Acknowledgments The authors would like to thank Professor Júlio S. Rebouças for his valuable suggestions. We are also grateful to the CNPq (476201/2009-8, 303856/2008-4, 303860/2008-1, 570565/2008-1) and CAPES (Brazilian agencies) for financial support.

References

- Loeppky RN, Michejda CJ (eds) (1994) Nitrosamines and related N-nitroso compounds: chemistry and biochemistry ACS symposium series 553. American Chemical Society, Washington DC
- Lijinsky W (1992) Chemistry and biology of N-Nitroso compounds. Cambridge University Press, Cambridge
- Preussmann R, Stewart BW, Searle CE (eds) (1984) Chemical carcinogens vol 2, 2nd edn. ACS monograph 182, Washington DC
- Loeppky RN, Outram JR (1982) In: Nitroso compounds: occurrence and biological effects. IARC, Lyon
- IARC (1978) Some N-nitroso compounds, IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans, vol 17. IARC, Lyon
- Daniel P, Arenas JF, Otero JC, Soto J (2007) J Org Chem 72:4741
- Wanno B, Ruangpornvisuti V (2006) J Mol Struct THEOCHEM 766:159
- Miura M, Sakamoto S, Yamaguchi K, Ohwada T (2000) Tetrahedron Lett 41:3637
- Ohwada T, Miura M, Tanaka H, Sakamoto S, Yamaguchi K, Ikeda H, Inagaki S (2001) J Am Chem Soc 123:10164
- Cho K-H (2005) Bull Korean Chem Soc 26:32
- Sarker H, Greer ML, Blackstock SC (1996) J Org Chem 61:3177
- Looney CE, Phillips WD, Reilly EL (1957) J Am Chem Soc 79:6136
- Krebs B, Mandt J (1975) Chem Ber 108:1130
- Rademacher P, Stølevik R (1969) Acta Chem Scand 23:660
- Geun-Bae Yi, Khan MA, Richter-Addo GB (1995) J Am Chem Soc 117:7850
- Chen Li Yi, Geun-Bae WL-S, Dharmawardana UR, Dart AC, Khan MA, Richter-Addo GB (1998) Inorg Chem 37:4677
- Lee J, Chen Li, West AH, Richter-Addo GB (2002) Chem Rev 102:1019
- Wang L-S Chen, Khan MA, Richter-Addo GB (1996) Chem Commun 323
- Jiang P, Qian X, Li C, Qiao C, Wang D (1997) Chem Phys Lett 277:508
- Becke AD (1993) J Chem Phys 98:5648
- Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1994) J Phys Chem 98:11623
- Møller C, Plesset MS (1934) Phys Rev 46:618
- Bartlett RJ, Silver DM (1975) J Chem Phys 62:3258
- Dunning TH Jr (1989) J Chem Phys 90:1007
- Do Monte SA, Ventura E, da Costa TF, Santana SR (2011) Struct Chem, in press
- Glendening ED, Reed AE, Carpenter JEF, Weinhold F NBO Version 31
- Martin F, Zipse H (2005) J Comput Chem 26:97
- Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA et al (2004) Gaussian 03, Revision E01. Gaussian Inc, Wallingford
- Sadlej-Sosnowska N, Krygowski TM (2009) Chem Phys Lett 476:191
- <http://cccbdbnist.gov/vibscalejustasp>

Gas adsorption on the Zn-, Pd- and Os-doped armchair (5,5) single-walled carbon nanotubes

Chanukorn Tabtimsai · Somchai Keawwangchai ·
Banchob Wannoo · Vithaya Ruangpornvisuti

Received: 11 September 2010 / Accepted: 17 March 2011 / Published online: 28 April 2011
© Springer-Verlag 2011

Abstract The adsorption of NO₂, NH₃, H₂O, CO₂ and H₂ gases on the undoped, Zn-, Pd- and Os-doped armchair (5,5) single-walled carbon nanotubes (SWCNTs) were studied using density functional method. The adsorptions of these five gases on the Zn-, Pd- and Os-doped SWCNTs are obviously stronger than on the undoped SWCNT and their adsorption abilities are in the same order: NO₂ > NH₃ > H₂O > CO₂ > H₂. Adsorption energies for all the studied gases on the undoped, Zn-, Pd- and Os-doped SWCNTs computed at the B3LYP/LanL2DZ level are reported.

Keywords Adsorption · Carbon nanotubes · Gases · Osmium-doped · Palladium-doped · Zinc-doped

Introduction

Carbon nanotubes were discovered by Iijima [1] and single-walled carbon nanotubes (SWCNTs) were found [2]. SWCNTs have generated strong interests because of their noticeable chemical and physical properties and potential applications in mechanical, electrical were widely studied [3–7]. Adsorptions of NO₂ [8–11], N₂O₄ [10], O₂ [9], NH₃ [9, 12], N₂ [9], CO₂ [9, 11], CH₄ [9],

H₂O [9], H₂ [9, 13] and Ar [9] gases on unloaded-SWCNTs were studied. Hydrogen storage by adsorption on SWCNTs and in multiple-walled carbon nanotubes (MWCNTs) was reviewed [14]. Due to electronic properties, NO₂ and O₂ adsorption are sensitive to the electronic properties of SWCNTs [9]. As no charge transfer between the tube and molecules, NO₂ and CO₂ molecules adsorb weakly on the tube wall [11]. Our previous works, adsorption of water [15], C1–C3 alkoxides [16] and nitrosamine [17] on various cap-ended and open-ended armchair (5,5) single-walled carbon nanotubes were studied. The sidewall additions of diazomethane to (n,n), n=3–10 armchair SWCNTs on two different orientations of C–C bonds in our previous work [18] were studied using ONIOM approach.

The metal-doped SWCNTs on their adsorption of small gaseous molecules were therefore widely studied for their gas sensing properties. The adsorptions of CO [19, 20], NO [19], NH₃ [19], N₂ [19], H₂ [19, 21], C₂H₄ [19, 21] and C₂H₂ [19] gases on Pt-doped SWCNTs were studied. The adsorption ability of CO [20], H₂ [21], C₂H₄ [21] onto the SWCNTs can be enhanced by Pt doping into the side wall. The CO-adsorption on Rh-decorated (8,0) SWCNT [22] was studied using density functional theory (DFT) calculations. The adsorption of H₂ on the Pd-SWCNT [23]. The experiments have also shown that transition metals (TMs) nanoparticles of Co, Ni, Ir, Pt, Ag, Rh, and Pd dispersed in carbon nanostructure have been synthesized and measured their hydrogen storage [24]. Moreover, theoretical calculations show the strong interaction between TMs (Sc, Ti, V, Ce, Mn, Fe, Co and Ni) with SWCNTs [25–27]. These results are described by the electron transferring process between gases and tubes resulting in the increasing of the conductance of the nanotubes [19, 28]. Recently, laboratory and theoretical studies have shown that SWCNT doped Rh nanoparticle can be used to detect CO at room temperature [20]. Adsorption of NH₃ on Stone–Wales defective carbon

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1047-y) contains supplementary material, which is available to authorized users.

C. Tabtimsai · S. Keawwangchai · B. Wannoo
Center of Excellence for Innovation in Chemistry and Department
of Chemistry, Faculty of Science, Mahasarakham University,
Mahasarakham 44150, Thailand

V. Ruangpornvisuti (✉)
Department of Chemistry, Faculty of Science,
Chulalongkorn University,
Bangkok 10330, Thailand
e-mail: vithaya.r@chula.ac.th

annotated was investigated using density functional theory [29]. Moreover, SWCNTs coated with Pd nanoparticles as the sensors have high sensitivity to H₂ molecules [30].

The Al-doped SWCNT presents high sensitivity to CO, compared with the intrinsic SWCNT, as indicated by the calculated geometrical structures and electronic properties for these systems [31]. B or N doping into the side wall of SWCNTs can increase the detection sensitivity for CO and H₂O molecules compared to pristine SWCNTs [32] and B doping onto SWCNTs also makes it highly sensitive for HNC and HCHO adsorptions [33]. The adsorption of NH₃ and NO₂ on B- and N-doped SWCNTs was investigated by using density functional computations to exploit their potential applications as gas sensors [34]. The calculated results indicate that NH₃ and NO₂ show the physisorption on metal doped sites with weakly binding energy and slight charge transfers [34].

In this work, adsorption of gaseous NO₂, NH₃, H₂O, CO₂ and H₂ on undoped armchair (5,5) SWCNT and Zn-, Pd- and Os-doped SWCNTs have been investigated using density functional method. Electronic properties for all adsorption complexes of all the studied gases on the undoped, Zn-, Pd- and Os-doped armchair (5,5) SWCNTs have been determined and their structures have been reported.

Computational methodology

Geometrical optimizations of SWCNT with and without doped metals and their gas adsorption structures were carried out using restricted density functional theory (DFT)

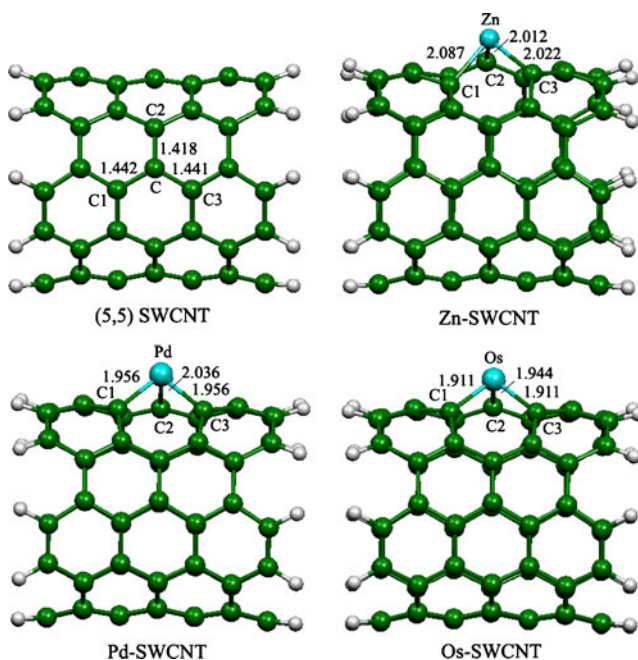


Fig. 1 The B3LYP/LanL2DZ optimized structures of the pristine armchair (5,5) SWCNT and Zn-, Pd- and Os-doped SWCNTs

approach. Seven-layered (5,5) armchair SWCNT was applied for all models. The hydrogen atoms were used to saturate the carbon atoms with dangling bonds at the two ends. The calculations have been performed with hybrid density functional B3LYP, the Becke's three-parameter exchange functional [35] with the Lee–Yang–Parr correlation functional [36], using the Los Alamos LanL2DZ split-valence basis set [37–39]. All calculations were performed using GAUSSIAN 03 program [40]. The electronic densities of states (DOSs) for the SWCNTs doped Zn, Pd and Os atoms and their gas adsorption were investigated and plotted by the GaussSum 2.1.4 program [41].

Table 1 The orbital energies, E_{LUMO} , E_{HOMO} and frontier molecular orbital energy gap, ΔE_{H-L} for adsorption of gas on Zn-, Pd- and Os-doped SWCNTs, and change of ΔE_{H-L} ($\Delta\Delta E_{H-L}$) during adsorption, computed at the B3LYP/LanL2DZ level

Adsorbent/adsorption configuration	E_{LUMO}^a	E_{HOMO}^a	ΔE_{H-L}^a	$\Delta\Delta E_{H-L}^{a,b}$
SWCNT system:				
SWCNT	-2.748	-4.762	2.014	
NO ₂ /SWCNT	-2.966	-4.925	1.959	-0.05
NH ₃ /SWCNT	-2.694	-4.925	2.231	0.22
H ₂ O/SWCNT	-2.748	-4.953	2.204	0.19
CO ₂ /SWCNT	-2.640	-4.844	2.204	0.19
H ₂ /SWCNT	-2.667	-4.871	2.204	0.19
Zn-SWCNT system:				
Zn-SWCNT	-3.293	-4.680	1.388	
NO ₂ /Zn-SWCNT	-3.973	-5.170	1.197	-0.19
NH ₃ /Zn-SWCNT	-2.830	-4.408	1.578	0.19
H ₂ O/Zn-SWCNT	-2.912	-4.490	1.578	0.19
CO ₂ /Zn-SWCNT	-3.129	-4.544	1.415	0.03
H ₂ /Zn-SWCNT	-3.293	-4.626	1.333	-0.05
Pd-SWCNT system:				
Pd-SWCNT	-2.857	-4.463	1.606	
NO ₂ /Pd-SWCNT	-3.048	-4.980	1.932	0.33
NH ₃ /Pd-SWCNT	-4.218	-2.558	1.660	0.05
H ₂ O/Pd-SWCNT	-4.218	-2.558	1.660	0.05
CO ₂ /Pd-SWCNT	-2.748	-4.381	1.633	0.03
H ₂ /Pd-SWCNT	-4.436	-2.803	1.633	0.03
Os-SWCNT system:				
Os-SWCNT	-3.021	-4.599	1.578	
NO ₂ /Os-SWCNT	-3.020	-4.925	1.905	0.33
NH ₃ /Os-SWCNT	-2.286	-4.299	2.014	0.44
H ₂ O/Os-SWCNT	-2.286	-4.299	2.014	0.44
CO ₂ /Os-SWCNT	-2.558	-4.490	1.932	0.35
H ₂ /Os-SWCNT	-2.857	-4.544	1.687	0.11

^a In eV

^b Defined as $\Delta E_{H-L}(\text{adsorbed SWCNT}) - \Delta E_{H-L}(\text{SWCNT})$ and $\Delta E_{H-L}(\text{adsorbed M-doped SWCNT}) - \Delta E_{H-L}(\text{M-doped SWCNT})$ for systems of SWCNT and metal-doped SWCNTs, respectively

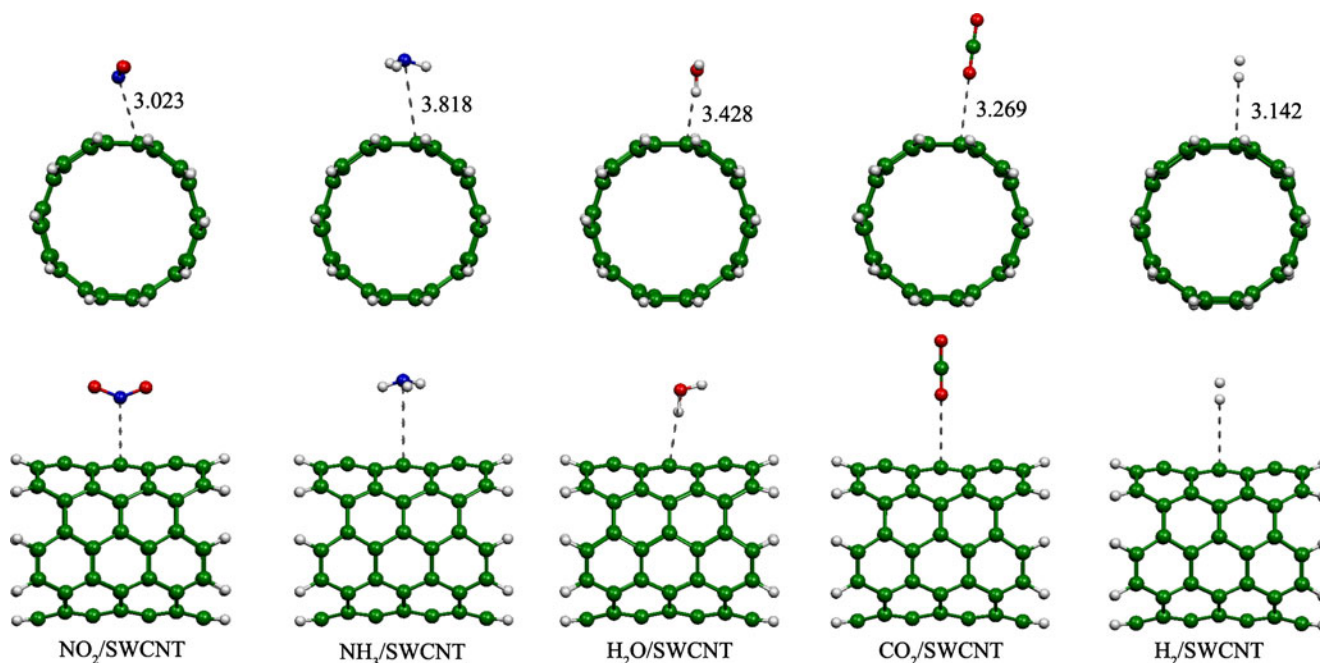


Fig. 2 The B3LYP/LanL2DZ optimized structures of various gases adsorbed on the pristine (5,5) SWCNT. Top and bottom are front and side views of tubes. Bond distances are in Å

The SWCNT was doped with Zn, Pd and Os atoms whereas the small molecular gaseous NO₂, NH₃, H₂O, CO₂ and H₂ were chosen for the gas adsorption study of SWCNT with and without doped metals. The adsorption of these small molecular gaseous have been studied on the ground-state surfaces of the Zn-, Pd- and Os-doped SWCNTs using the B3LYP/LanL2DZ computations. As the adsorptions of NO₂ on the metal-doped SWCNTs were

found to be the strongest energies, adsorption energies of NO₂ adsorbed on the spin-polarized (high spin) surfaces of Zn-, Pd- and Os-doped SWCNTs using the unrestricted B3LYP/LanL2DZ computations were also examined. It was found that the adsorption energies of NO₂ on the spin-polarized Zn-, Pd- and Os-doped SWCNTs are smaller than on the ground state Zn-, Pd- and Os-doped SWCNTs but still high, as shown in Table S1, the Electronic

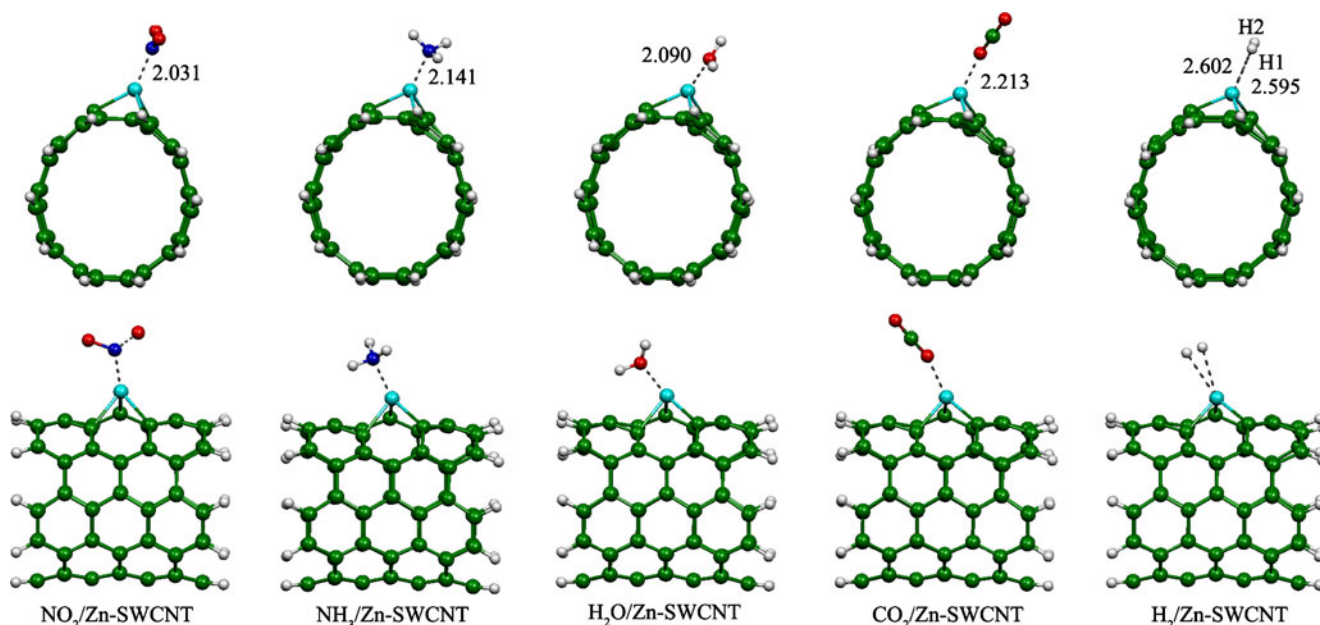


Fig. 3 The B3LYP/LanL2DZ optimized structures of various gases adsorbed on the Zn-doped SWCNT. Top and bottom are front and side views of tubes. Bond distances are in Å

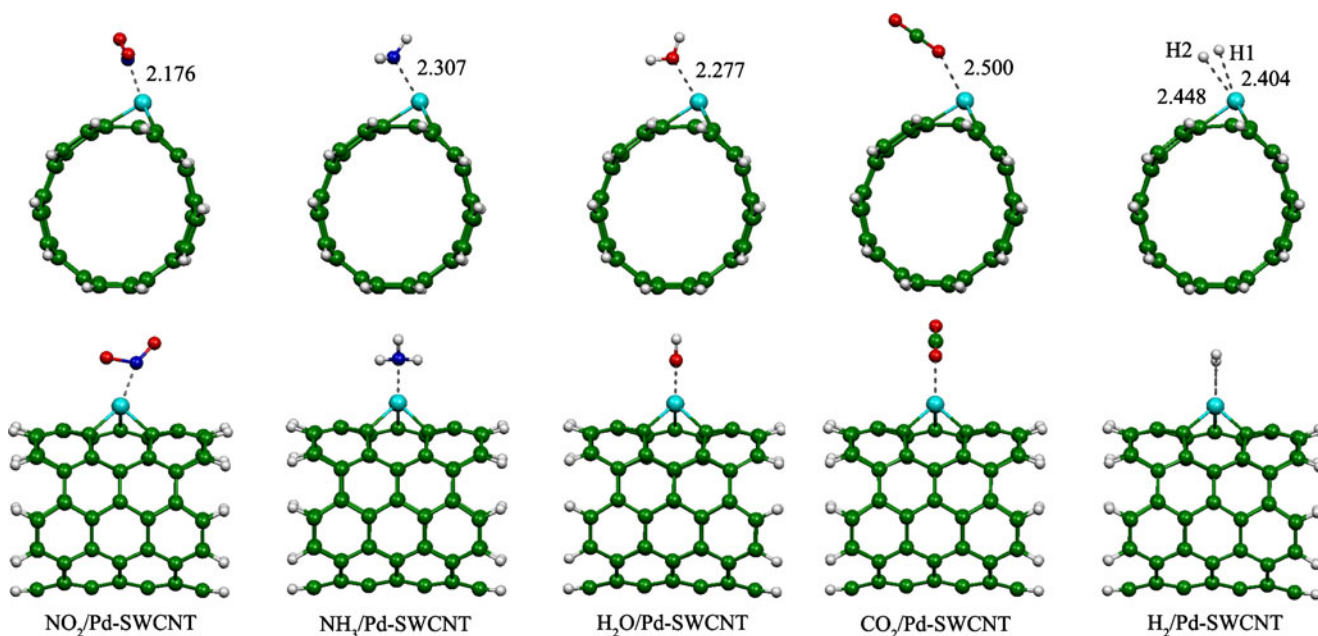


Fig. 4 The B3LYP/LanL2DZ optimized structures of various gases adsorbed on the Pd-doped SWCNT. Top and bottom are front and side views of tubes. Bond distances are in Å

supplementary material (ESM). The adsorption energies of NO₂ on ground-state and spin-polarized surfaces of the Zn-, Pd- and Os-doped SWCNTs are still high although they are not in the same order. Nevertheless, the adsorptions of these small molecular gaseous on the ground-state surfaces of the Zn-, Pd- and Os-doped SWCNTs are good enough to describe adsorption abilities of the ground-state Zn-, Pd- and Os-doped SWCNTs.

Results and discussion

Properties of metal-doped armchair (5,5) SWCNTs

The B3LYP/LanL2DZ-optimized structures of the undoped, Zn-, Pd- and Os-doped armchair (5,5) SWCNTs are shown in Fig. 1. It seems to show that metal atom of metal-doped SWCNT is located between the top of three

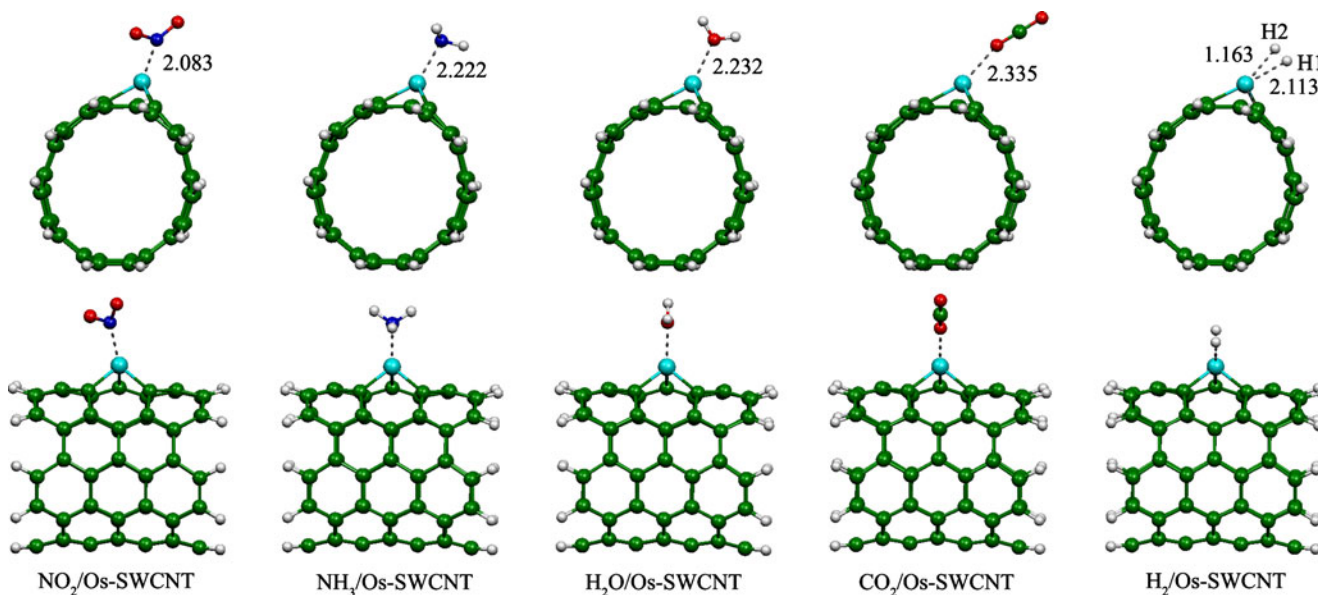


Fig. 5 The B3LYP/LanL2DZ optimized structures of various gases adsorbed on the Os-doped SWCNT. Top and bottom are front and side views of tubes. Bond distances are in Å

neighboring carbon atoms of SWCNT. The bond lengths between doped metal and carbon atom, M–C are in order: Zn–C > Pd–C > Os–C. These M–C bond lengths correspond with sizes of Zn, Pd and Os metals.

The computed IR spectra of the B3LYP/LanL2DZ-optimized structures of the undoped, Zn-, Pd- and Os-doped SWCNTs are shown in Fig. S1. It shows that intensities of vibration frequencies within the range of 1100 to 1600 cm⁻¹ of metal-doped SWCNTs are remarkably increased as compared with the undoped SWCNT. The principal vibration frequencies and their corresponding intensities are listed in Table S2. There are two dominant spectra of which vibration intensities are very strong namely $\delta_{\text{CCH}} \approx 800 \text{ cm}^{-1}$ and $\nu_{\text{CH}} (\text{sym.}) \approx 3100 \text{ cm}^{-1}$. The lowest unoccupied molecular orbital (E_{LUMO}) energy, the highest occupied molecular orbital (E_{HOMO}) energy and frontier molecular orbital energy gap ($\Delta E_{\text{H-L}}$) for the undoped, Zn-, Pd- and Os-doped SWCNTs are shown in Table 1. The relative reactivities of these SWCNTs are in

order: Zn-doped > Os-doped > Pd-doped > undoped SWCNTs.

Adsorption of gases on the undoped and metal-doped armchair (5,5) SWCNTs

The B3LYP/LanL2DZ-optimized structures of adsorption of gaseous NO₂, NH₃, H₂O, CO₂ and H₂ on the undoped, Zn-, Pd- and Os-doped armchair (5,5) SWCNTs are shown in Figs. 2, 3, 4 and 5, respectively. The E_{LUMO} , E_{HOMO} and $\Delta E_{\text{H-L}}$ for adsorption complexes of the undoped, Zn-, Pd- and Os-doped SWCNTs, listed in Table 1 show that the relative reactivities of these complexes are in orders: NO₂/SWCNT > H₂O/SWCNT \approx CO₂/SWCNT \approx H₂/SWCNT > NH₃/SWCNT, NO₂/Zn-SWCNT > H₂/Zn-SWCNT > CO₂/Zn-SWCNT > NH₃/Zn-SWCNT \approx H₂O/Zn-SWCNT, CO₂/Pd-SWCNT \approx H₂/Pd-SWCNT \approx NH₃/Pd-SWCNT > H₂O/Pd-SWCNT > NO₂/Pd-SWCNT and H₂/Os-SWCNT > NO₂/Os-SWCNT

Table 2 The selected geometrical parameters for the undoped, Zn-, Pd- and Os-doped SWCNTs and their adsorption complexes with studied gases, computed at the B3LYP/LanL2DZ level

Adsorbent/adsorption configuration	r^a				θ^b		
	C1-M ^{c,d}	C2-M ^{c,d}	C3-M ^{c,d}	M'-X ^e	C1-M-C2 ^{c,d}	C2-M-C3 ^{c,d}	C3-M-C1 ^{c,d}
SWCNT	1.442	1.418	1.441		118.6	118.6	120.3
NO ₂ /SWCNT	1.442	1.422	1.441	3.023 (C–N)	118.5	118.5	120.3
NH ₃ /SWCNT	1.441	1.417	1.441	3.818 (C–N)	118.6	118.6	120.2
H ₂ O/SWCNT	1.444	1.419	1.442	3.428 (C–O)	118.5	118.5	120.2
CO ₂ /SWCNT	1.441	1.417	1.441	3.269 (C–O)	118.6	118.6	120.2
H ₂ /SWCNT	1.441	1.418	1.441	3.142 (C–H)	118.7	118.7	120.2
Zn-SWCNT	2.087	2.012	2.022		82.7	100.9	85.8
NO ₂ /Zn-SWCNT	2.226	2.060	2.138	2.031 (Zn–N)	82.0	89.4	75.3
NH ₃ /Zn-SWCNT	2.151	2.040	2.057	2.141 (Zn–N)	78.9	96.2	81.8
H ₂ O/Zn-SWCNT	2.148	2.036	2.048	2.090 (Zn–O)	78.5	96.8	82.6
CO ₂ /Zn-SWCNT	2.102	2.018	2.029	2.213 (Zn–O)	88.9	99.5	85.5
H ₂ /Zn-SWCNT	2.075	2.009	2.018	2.595 (Zn–H1), 2.602 (Zn–H2)	89.7	101.3	87.4
Pd-SWCNT	1.956	2.036	1.956		85.7	85.7	93.8
NO ₂ /Pd-SWCNT	1.987	2.013	1.982	2.176 (Pd–N)	85.8	85.5	89.9
NH ₃ /Pd-SWCNT	1.976	2.042	1.976	2.307 (Pd–N)	85.1	85.1	88.0
H ₂ O/Pd-SWCNT	1.967	2.042	1.967	2.277 (Pd–O)	85.4	85.4	89.6
CO ₂ /Pd-SWCNT	1.962	2.043	1.962	2.500 (Pd–O)	85.2	85.2	91.7
H ₂ /Pd-SWCNT	1.961	2.053	1.961	2.404 (Pd–H1), 2.448 (Pd–H2)	84.9	84.9	92.0
Os-SWCNT	1.911	1.944	1.911		90.6	90.6	97.6
NO ₂ /Os-SWCNT	1.907	1.993	1.899	2.083 (Os–N)	87.9	87.8	94.21
NH ₃ /Os-SWCNT	1.906	1.983	1.906	2.222 (Os–N)	88.2	88.2	96.5
H ₂ O/Os-SWCNT	1.905	1.972	2.000	2.232 (Os–O)	88.2	88.2	96.8
CO ₂ /Os-SWCNT	1.911	1.957	1.911	2.335 (Os–O)	89.3	89.3	97.1
H ₂ /Os-SWCNT	1.915	1.963	1.915	2.113 (Os–H1), 2.163 (Os–H2)	89.7	89.7	97.3

^a In Å

^b In degree

^c C1, C2 and C3 are atoms on the SWCNT which are defined in Fig. 1

^d M is atom of Zn, Pd or Os metal which is doped on SWCNT, see Fig. 1

^e M' is atom of Zn, Pd or Os metal which is doped on SWCNT or C atom which is labeled in Fig. 1 and X is atom of adsorbed gas, see Fig. 1

> CO₂/Os–SWCNT > NH₃/Os–SWCNT ≈ H₂O/Os–SWCNT for the systems of undoped, Zn–, Pd– and Os–doped SWCNTs, respectively. Table 1 also shows that reactivities of all adsorbents are increased during adsorption except adsorption of NO₂ on the undoped SWCNT and Zn–doped SWCNT and adsorption of H₂ on Zn–doped SWCNT.

The selected geometrical parameters for on the undoped, Zn–, Pd– and Os–doped SWCNTs and their adsorption complexes with the studied gases are listed in Table 2. It was found that the adsorption of gas on the carbon center (C, see Fig. 1) of undoped SWCNT has hardly ever affected the bond lengths C1–C, C3–C (see Fig. 1 and Table S3). For the bond distances between the closest atom of gases and metal center of metal–doped SWCNTs are in increasing orders: Zn–N(NO₂) > Zn–O(H₂O) > Zn–N(NH₃) > Zn–O(CO₂) > Zn–H(H₂), Pd–N(NO₂) > Pd–O(H₂O) > Pd–N(NH₃) > Pd–H(H₂) > Pd–O(CO₂) and Os–N(NO₂) > Os–H(H₂) > Os–N(NH₃) > Os–O(H₂O) > Os–O(CO₂) for Zn–, Pd– and Os–doped SWCNTs, respectively.

The adsorption energy (ΔE_{ads}) of gas adsorbed on the undoped, Zn–, Pd– and Os–doped SWCNTs and partial charge transfer (PCT) [42] during adsorption are shown in Table 3. It was found that PCT for all adsorbed gases on the undoped, Zn–, Pd– and Os–doped SWCNTs are positive except the adsorption of NO₂ on the undoped, Zn–, Pd– and Os–doped SWCNTs being negative PCT values which is implied that the direction of PCT found in NO₂ adsorption is different to the other gas adsorptions. The selected NBO charges for adsorption of gases on the undoped, Zn–, Pd– and Os–doped SWCNTs are shown in Table S4. The adsorption energies of gases adsorbed on the undoped, Zn–, Pd– and Os–doped SWCNTs are given in Table 3 and their plot are shown in Fig. 6. The adsorptions of all studied gases on the undoped SWCNT are weak and their abilities are in decreasing order: H₂O ($\Delta E_{\text{ads}}=-8.39$ kJ mol⁻¹), > NO₂ ($\Delta E_{\text{ads}}=-5.58$ kJ mol⁻¹) > NH₃ ($\Delta E_{\text{ads}}=-03.19$ kJ mol⁻¹) > CO₂ ($\Delta E_{\text{ads}}=-1.61$ kJ mol⁻¹) > H₂ ($\Delta E_{\text{ads}}=-0.66$ kJ mol⁻¹). The adsorption abilities of gases on the Zn–, Pd– and Os–doped SWCNTs are remarkably increased as compared with the undoped SWCNT. The adsorptions of all studied gases on the Zn–, Pd– and Os–doped SWCNTs are obviously stronger than on the undoped SWCNT and their adsorption abilities are in the same order: NO₂ > NH₃ > H₂O > CO₂ > H₂.

Nevertheless, the adsorption abilities of all the studied gases on Zn–, Pd– and Os–doped SWCNTs are in the same order. The adsorption of the NO₂ on the Zn–, Pd– and Os–doped SWCNTs which are the highest strengths are much higher than of the NH₃, the second highest strengths. The adsorption ability order for the NO₂ adsorption on the

Table 3 Adsorption energy (ΔE_{ads}) and partial charge transfer (PCT) of gas adsorbed on the undoped, Zn–, Pd– and Os–doped SWCNTs, computed at the B3LYP/LanL2DZ level

Adsorbent/adsorption configuration	ΔE_{ads} ^a	PCT ^b
SWCNT:		
NO ₂ /SWCNT	-5.58	-0.035
NH ₃ /SWCNT	-3.19	0.005
H ₂ O/SWCNT	-8.39	0.002
CO ₂ /SWCNT	-1.61	0.002
H ₂ /SWCNT	-0.66	0.000
Zn–SWCNT:		
NO ₂ /Zn–SWCNT	-194.57	-0.798
NH ₃ /Zn–SWCNT	-123.61	0.074
H ₂ O/Zn–SWCNT	-105.15	0.048
CO ₂ /Zn–SWCNT	-39.98	0.020
H ₂ /Zn–SWCNT	-5.61	0.016
Pd–SWCNT:		
NO ₂ /Pd–SWCNT	-201.97	-0.626
NH ₃ /Pd–SWCNT	-100.97	0.119
H ₂ O/Pd–SWCNT	-79.70	0.070
CO ₂ /Pd–SWCNT	-26.46	0.024
H ₂ /Pd–SWCNT	-8.74	0.018
Os–SWCNT:		
NO ₂ /Os–SWCNT	-240.99	-0.568
NH ₃ /Os–SWCNT	-155.97	0.194
H ₂ O/Os–SWCNT	-108.79	0.127
CO ₂ /Os–SWCNT	-45.63	0.051
H ₂ /Os–SWCNT	-24.86	0.187

^a In kJ mol⁻¹

^b Defined as a change of adsorbate's charges during adsorption, in e

metal–doped SWCNTs is Os–doped ($\Delta E_{\text{ads}}=-240.99$ kJ mol⁻¹) >> Pd–doped ($\Delta E_{\text{ads}}=-201.97$ kJ mol⁻¹) > Zn–doped ($\Delta E_{\text{ads}}=-194.57$ kJ mol⁻¹) SWCNTs. These data imply that the Os–doped SWCNT can be circumstantially used as a sensor for nitrogen dioxide gas detection.

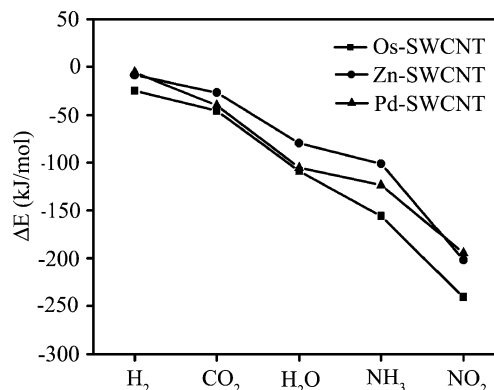


Fig. 6 Plot of adsorption energies (ΔE_{ads}) of the various gases adsorbed on the Zn–, Pd– and Os–doped (5,5) SWCNTs

The plots of the HOMOs and LUMOs of the adsorption complexes of all the studied gases with the Zn-, Pd- and Os-doped SWCNTs, are shown in Figs. S2, S3 and S4, in the ESM, respectively. The plots show that all the HOMOs of gas adsorption complexes with the Pd- and Os-doped SWCNTs are delocalized over the metal atoms but the HOMOs of gas adsorption complexes with the Zn-doped SWCNT are located over carbon atoms around the Zn atom. This implies that the Pd and Os atoms enhance gas adsorption abilities of the Pd- and Os-doped SWCNTs, respectively.

The density of states (DOSs) determined for the Zn-, Pd- and Os-doped SWCNTs and their adsorption complexes with gas are shown in Figs. S5, S6 and S7, in the ESM, respectively. These figures show that DOSs peaks of the spin up and spin down for the bare adsorbents and their gas adsorption are all symmetries except for the NO₂ adsorbed on the Zn-, Pd- and Os-doped SWCNTs are asymmetries. Asymmetric peaks of the spin up and spin down for the NO₂ adsorbed on the Pd- and Os-doped SWCNTs are remarkably different from the Zn-doped SWCNT. Nevertheless, the DOS plots for the gases adsorbed on the Zn-, Pd- and Os-doped SWCNTs are slightly different from their corresponding bare metal-doped SWCNTs.

Conclusions

The optimized structures of the undoped armchair (5,5) SWCN, Zn-, Pd- and Os-doped armchair (5,5) SWCNTs and their adsorption of NO₂, NH₃, H₂O, CO₂ and H₂ were obtained using the B3LYP/LanL2DZ calculations. The closest C–C bonds of the undoped and metal-doped SWCNTs at adsorption area are in order: Zn-doped > Pd-doped > Os-doped >> undoped SWCNTs. The bond lengths between doped metal and carbon atom, M–C are in order: Zn–C > Pd–C > Os–C. The computed IR spectra of the B3LYP/LanL2DZ-optimized structures of the undoped, Zn-, Pd- and Os-doped SWCNTs were analyzed. The relative reactivities of these adsorbents are in order: Zn-doped > Os-doped > Pd-doped > undoped SWCNTs.

The relative reactivities of these complexes are in orders: NO₂/SWCNT > H₂O/SWCNT ≈ CO₂/SWCNT ≈ H₂/SWCNT > NH₃/SWCNT, NO₂/Zn-SWCNT > H₂/Zn-SWCNT > CO₂/Zn-SWCNT > NH₃/Zn-SWCNT ≈ H₂O/Zn-SWCNT, CO₂/Pd-SWCNT ≈ H₂/Pd-SWCNT ≈ NH₃/Pd-SWCNT > H₂O/Pd-SWCNT > NO₂/Pd-SWCNT and H₂/Os-SWCNT > NO₂/Os-SWCNT > CO₂/Os-SWCNT > NH₃/Os-SWCNT ≈ H₂O/Os-SWCNT for the systems of undoped, Zn-, Pd- and Os-doped SWCNTs, respectively. The adsorption abilities of all the studied gases on the undoped, Zn-, Pd- and Os-doped SWCNTs are in orders: H₂O > CO₂ > NO₂ > NH₃ > H₂, NO₂ >> NH₃ > H₂O >>

CO₂ >> H₂, NO₂ >> NH₃ > H₂O >> CO₂ >> H₂ and NO₂ >> NH₃ >> H₂O >> CO₂ >> H₂, respectively.

Acknowledgments The financial support (Grant No. MRG5180141) to BW by the Thailand Research Fund is gratefully acknowledged. Financial support from the Center for Innovation in Chemistry (PERCH-CIC), Commission on Higher Education, Ministry of Education is gratefully acknowledged. We also thank the Supramolecular Chemistry Research Unit, Department of Chemistry, Faculty of Science, Mahasarakham University, the Thailand Research Fund (TRF), the National Nanotechnology Center (NANOTEC), National Science and Technology Development Agency, for providing facility. The Center for Petroleum, Petrochemicals and Advanced Materials, Chulalongkorn University, Bangkok, Thailand is also acknowledged for partly financial support.

References

- Iijima S (1991) Nature 354:56–58
- Iijima S, Ichihashi T (1993) Nature 363:603–605
- Burchell TD (1996) Carbon materials for advances technologies. Elsevier Science Ltd, Kidlington
- Meyyappan M (2005) Carbon nanotubes: science and applications. CRC, Boca Raton, FL
- Dresselhaus MS, Dresselhaus G, Eklund PC (1996) Science of fullerenes and carbon nanotubes. Academic, New York
- Gogotsi Y (2006) Carbon Nanomaterials. CRC, Boca Raton, FL
- Popov V, Lambin P (2006) Carbon Nanotubes. Springer, Berlin
- Peng S, Cho K, Qi P, Dai H (2004) Chem Phys Lett 387:271–276
- Zhao JJ, Buldum A, Han J, Lu JP (2002) Nanotechnol 13:195–200
- Yim WL, Gong XG, Liu ZF (2003) J Phys Chem B 107:9363–9369
- Santucci S, Picozzi S, Di Gregorio F, Lozzi L, Cantalini C, Valentini L, Kenny JM, Delley B (2003) J Chem Phys 119:10904–10910
- Bauschlicher WC, Ricca A (2003) Phys Rev B 70:115409–115416
- Han SS, Lee HM (2004) Carbon 42:2169–2177
- Darkrima FL, Malbrunota P, Tartaglia GP (2002) Int J Hydrogen Energy 27:193–202
- Wanbayor R, Ruangpornvisuti V (2007) Chem Phys Lett 441:127–131
- Wanbayor R, Ruangpornvisuti V (2008) Carbon 46:12–18
- Ruangpornvisuti V (2010) J Mol Model 16:1127–1138
- Wanno B, Du AJ, Ruangpornvisuti V, Smith SC (2007) Chem Phys Lett 436:218–223
- Yeung CS, Liu LV, Wang YA (2008) J Phys Chem C 112:7401–7411
- Yeung CS, Liu LV, Wang YA (2007) J Theor Comput Nanosci 4:1108–1119
- Tian W, Liu LV, Wang Y (2006) Phys Chem Chem Phys 8:3528–3539
- Zhoa JX, Ding YH (2008) Mater Chem Phys 110:411–416
- Kong J, Chapline MG, Dai H (2001) Adv Mater 13:1384–1386
- Yang YX, Singh RK, Webley PA (2008) Adsorption 14:265–274
- Yang CK, Zhao J, Lu JP (2004) Nano Lett 4:561–563
- Dag S, Durgun E, Ciraci S (2004) Phys Rev B 69:121407(R)
- Wang G, Huang Y (2009) J Phys Chem C 112:9128–9132
- An W, Turner H (2009) Chem Phys Lett 482:274–280
- Andzelm J, Govind N, Maiti A (2006) Chem Phys Lett 421:58–62
- Xiao H, Li SH, Cao JX (2009) Chem Phys Lett 483:111–114

31. Wang R, Zhang D, Sun W, Han Z, Liu C (2007) *J Mol Struct THEOCHEM* 806:93–97
32. Peng C, Cho K (2003) *Nano Lett* 3:513–517
33. Zhang Y, Zhang D, Liu C (2006) *J Phys Chem B* 110:4671–4674
34. Bai L, Zhou Z (2007) *Carbon* 45:2105–2110
35. Becke AD (1993) *J Chem Phys* 98:5648–5652
36. Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37:785–789
37. Hay PJ, Wadt WR (1985) *J Chem Phys* 82:270–283
38. Wadt WR, Hay PJ (1985) *J Chem Phys* 82:284–298
39. Hay PJ, Wadt WR (1985) *J Chem Phys* 82:299–310
40. Frisch MJ et al. (2004) *Gaussian03, Rev D.02*, Gaussian Inc, Wallingford, CT
41. O'Boyle NM, Tenderholt AL, Langner KM (2008) *J Comput Chem* 9:839–845
42. Marichev VA (2009) *Physicochem Eng Aspects* 348:28–34

Quantum chemical investigation of the thermal pyrolysis reactions of the carboxylic group in a brown coal model

Shengyu Liu · Zhiqiang Zhang · Huifang Wang

Received: 18 December 2010 / Accepted: 29 March 2011 / Published online: 3 May 2011
© Springer-Verlag 2011

Abstract Different reaction pathways of the carboxylic group in a brown coal model were investigated by applying density function quantum chemical theory, examining the possible cross-linking and decomposition reactions between the hydrogen bonded carboxylic group–carboxylic group and the carboxylic group–hydroxyl group during the thermal pyrolysis process. The results show that bimolecular dehydration and decarboxylation of hydrogen bonded carboxylic groups have distinctly lower activation barriers and therefore, proceed preferentially at low temperature. The esterification reaction between the hydrogen bonded carboxylic group and hydroxyl group, together with unimolecular decarboxylation of isolated single carboxylic groups were also possible at moderate temperature. Aryl–aryl coupling is thought to occur via radical pyrolysis and recombination at relatively high temperature.

Keywords Decarboxylation · Thermal pyrolysis · Cross-linking · Brown coal · Quantum chemistry

Introduction

An understanding of the thermal pyrolysis of brown coal is important because it is the initial process for coal upgrading, combustion, and gasification [1, 2]. The O/C ratio of brown coal is the highest among all coal ranks. Oxygen groups in coal can be categorized into carboxyl, hydroxyl, carbonyl and ether functionalities. The first two groups constitute nearly half of the oxygen in brown coal. The chemistry relevant to the pyrolysis of coals is extremely complicated. During the pyrolysis of coal, the breakup of the coal macromolecular network and the resulting product formation are controlled by the relative rates of bond breaking, cross-linking, and mass transport. Oxygen functional groups, especially with the high concentration of carboxylic groups present in brown coals, have been implicated as the major contributors to thermal pyrolysis reactions [3].

Decarboxylation of carboxylic groups is considered to play an important role in the pyrolysis of brown coal [4]. However, the reaction mechanisms by which decarboxylation occurs in brown coals are not known. It has been proposed that decarboxylation reactions are the major contributors to the cross-linking processes observed during the thermal processing of brown coal [5, 6]. However, this proposal has been brought into question by studies on the pyrolysis of aromatic carboxylic acids, in which decarboxylation leads to only a limited amount of cross-linking [4, 7]. It was also found that the decarboxylation of coal model compounds yielded only small amounts of aryl–aryl coupling products, which would be

S. Liu (✉) · Z. Zhang (✉)
College of Mining Engineering,
Taiyuan University of Technology,
Taiyuan 030024, People's Republic of China
email: Liusyutyut@126.com

Z. Zhang
email: donaldzhang@foxmail.com

H. Wang
Key Laboratory of Interface Science and Engineering in
Advanced Materials, Taiyuan University of Technology,
Ministry of Education,
Taiyuan 030024, People's Republic of China

H. Wang
College of Materials Science and Engineering,
Taiyuan University of Technology,
Taiyuan 030024, People's Republic of China

stable at $t < 400$ °C and would constitute a low-temperature cross-link [5]. These results suggest that cross-linking processes in brown coal may not be related directly to the decarboxylation process but may result indirectly from intermediates formed from the reactions of aromatic carboxylic groups [8]. In addition, it has also been shown that hydrogen-bonded carboxylic and hydroxyl groups in coal cross-link at an early stage of pyrolysis, significantly affecting the subsequent main pyrolysis reactions [9].

To gain a better understanding of the role played by decarboxylation in the thermal pyrolysis process of brown coal, several possible reactions involving carboxylic groups have been studied in model compounds using quantum chemical methods. Because of the complex, ill-defined molecular compositions and structures of coals, along with the multiplicity of reaction possibilities when these materials undergo thermal pyrolysis, it is difficult to simulate the thermal pyrolysis reaction of the whole coal macromolecule. Therefore, small organic compounds were chosen as model compounds to explore possible thermal pyrolysis reactions in brown coal. In this study, benzoic acid and phenol were chosen as the molecular models because aromatic carboxylic and hydroxyl groups are known to be prevalent in brown coal and determine the surface chemistry of the coal.

Theoretical calculations

The Gaussian'03 package of programs [10] was used for density functional theory (DFT) quantum chemical calculations. Becke's three parameter (B3) gradient-corrected exchange functional was used, and the non-local correlation was provided by the Lee, Yang and Parr (LYP) expressions. All electronic and vibrational properties were calculated using the 6-311 + G (d, p) basis set, which includes diffuse functions on heavy atoms and polarization on heavy and hydrogen atoms—a requirement necessary for calculations of charged-separated states. This is also required for diradical structures where some atoms bear very large electron densities. The unrestricted open shell methodology was used in radical reaction pathways [11]. The minimum energy structures were achieved by allowing all the geometrical parameters to vary independently. The transition states were optimized by using synchronous transit-guided quasi-Newton (STQN) method. This method, implemented by Schlegel and co-workers [12, 13], uses a quadratic synchronous transit approach to get closer to the quadratic region of the transition state and then uses a quasi-Newton or eigenvector-following algorithm to complete the optimization. The transition structures (TS) obtained were confirmed by frequency calculations. Subsequent intrinsic reaction coordinate (IRC) calculations

were performed to ensure that the computed transition states correlate with the minima [14]. All values are unscaled, and the energies are reported in kcal mol⁻¹.

Results and discussion

Brown coal contains mainly abundant aromatic carboxylic groups, which exist in the form of isolated single carboxylic groups and hydrogen-bonded complexes. As regard to hydrogen-bonded complexes, apart from hydrogen-bonded carboxyl group–carboxyl group complexes, carboxyl group–hydroxyl group complexes associated with the hydrogen bond in the coal matrix were also considered in this study. The following sections examine possible reaction schemes, which include unimolecular decarboxylation, bimolecular coupling, bimolecular decarboxylation and esterification reactions between carboxyl group and hydroxyl group, for the carboxylic group in the brown coal matrix.

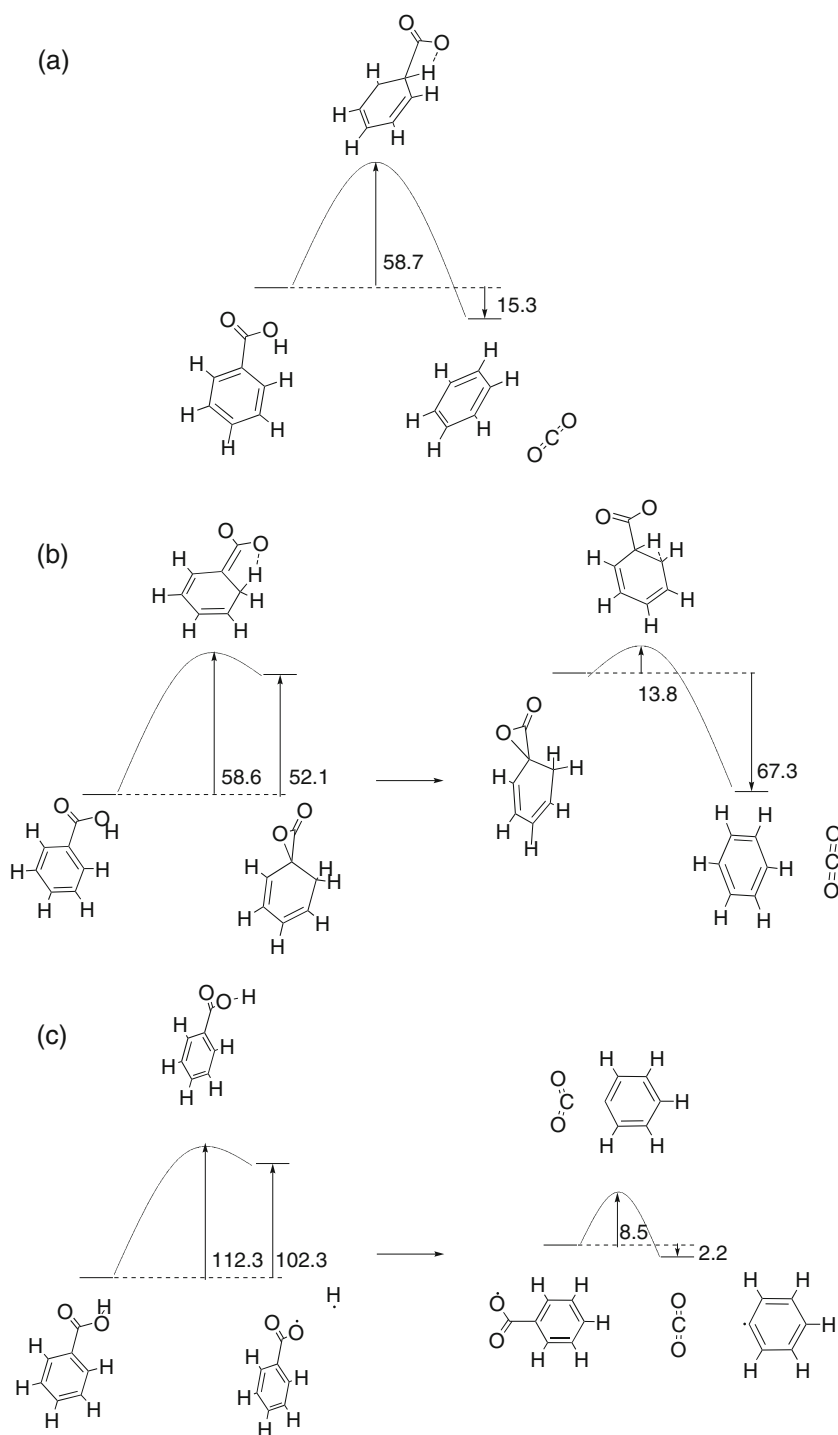
Unimolecular decarboxylation reaction

Three unimolecular decarboxylation mechanisms for benzoic acid are proposed and shown in Fig. 1. The first reaction pathway (Fig. 1a) is direct decarboxylation to form benzene and CO₂ from the reaction complex to the four-membered ring transition state. IRC studies show that this decarboxylation process starts with elongation of the C–C bond followed by hydrogen transfer towards the benzene ring. The energy required to attain the transition state geometry is 58.7 kcal mol⁻¹, which is in very good agreement with previous results (58.8 kcal mol⁻¹) reported by Li and Brill [15] at the level of B3LYP/6-31 + G(d,p).

The second possible reaction pathway for benzoic acid is the stepwise decarboxylation mechanism shown in Fig. 1b. The first step of this pathway is the migration of the hydrogen atom from the hydroxyl group to the ortho-C with a five-membered ring transition state. In the second step, the hydrogen atom returns to the ipso-C to form benzene, accompanied by the release of CO₂ from the benzene ring. The barrier height of the first step is 58.6 kcal mol⁻¹, which is higher than that of the second step (13.8 kcal mol⁻¹). Therefore, the rate-determined step is the first step in this reaction pathway.

The third possibility for the reaction mechanism is the stepwise radical process shown in Fig. 1c. The first step is the homolytic cleavage of benzoic acid, leading to benzoic and hydrogen radicals with activation energy of 112.3 kcal mol⁻¹. The second step is the decarboxylation of benzoic radical to form phenyl radical and CO₂ with an activation energy of 8.5 kcal mol⁻¹. From the values of the activation barrier, it is known that the rate-determining step is the first step.

Fig. 1 Energy diagrams of decarboxylation reactions for benzoic acid. **a** Concerted, **b** stepwise, **c** radical (energy scale in kcal mol⁻¹)



It can be noted that the first two reaction pathways have similar activation barriers, although the corresponding transition state structures are distinctly different. However, the activation barrier of the radical process is higher in comparison. Therefore, the possible involvement of a free-radical decarboxylation pathway can be ruled out from the unimolecular decarboxylation mechanism, which is consistent with the results of Eskay et al. [8].

Bimolecular coupling and the decarboxylation reaction

The next step in this computational study was to investigate the bimolecular coupling and decarboxylation reaction of the benzoic acid dimer. The bimolecular coupling is assumed to include four steps as shown in Fig. 2a. The first step of bimolecular coupling of benzoic acid is pyrolytic dehydration to its anhydride, which has been

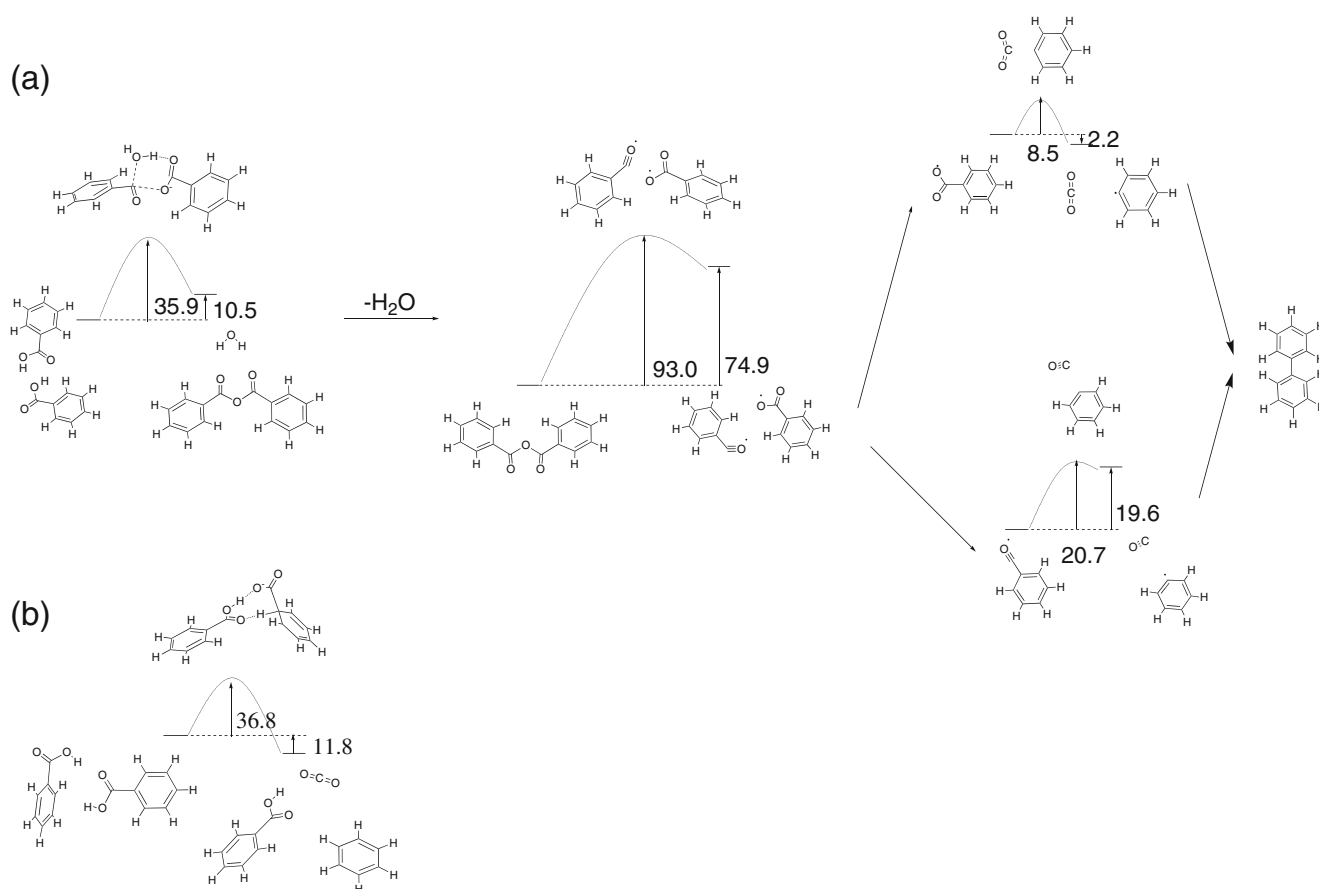


Fig. 2 Energy diagrams of the bimolecular reactions of benzoic acid. **a** Coupling, **b** decarboxylation (energy scale in kcal mol^{-1})

found in the pyrolysis of monocarboxylic acids [16]. The optimized structure of the reactant complex shows that two intermolecular hydrogen bonds are formed in the benzoic acid dimer. In the second stage, benzoic anhydride undergoes homolysis to form benzoyl and benzoyloxy radicals. Next, benzoyl and benzoyloxy radicals decarboxylate to form CO, CO₂ and two benzene radicals. Finally, these two aryl radicals form the cross-linking products without activation barrier. For this decarboxylation mechanism, the activation barrier of the second step ($93.0 \text{ kcal mol}^{-1}$), is higher than that of the first ($35.9 \text{ kcal mol}^{-1}$) and third (8.5 and $20.7 \text{ kcal mol}^{-1}$) steps and is therefore the rate-determining step of this reaction pathway.

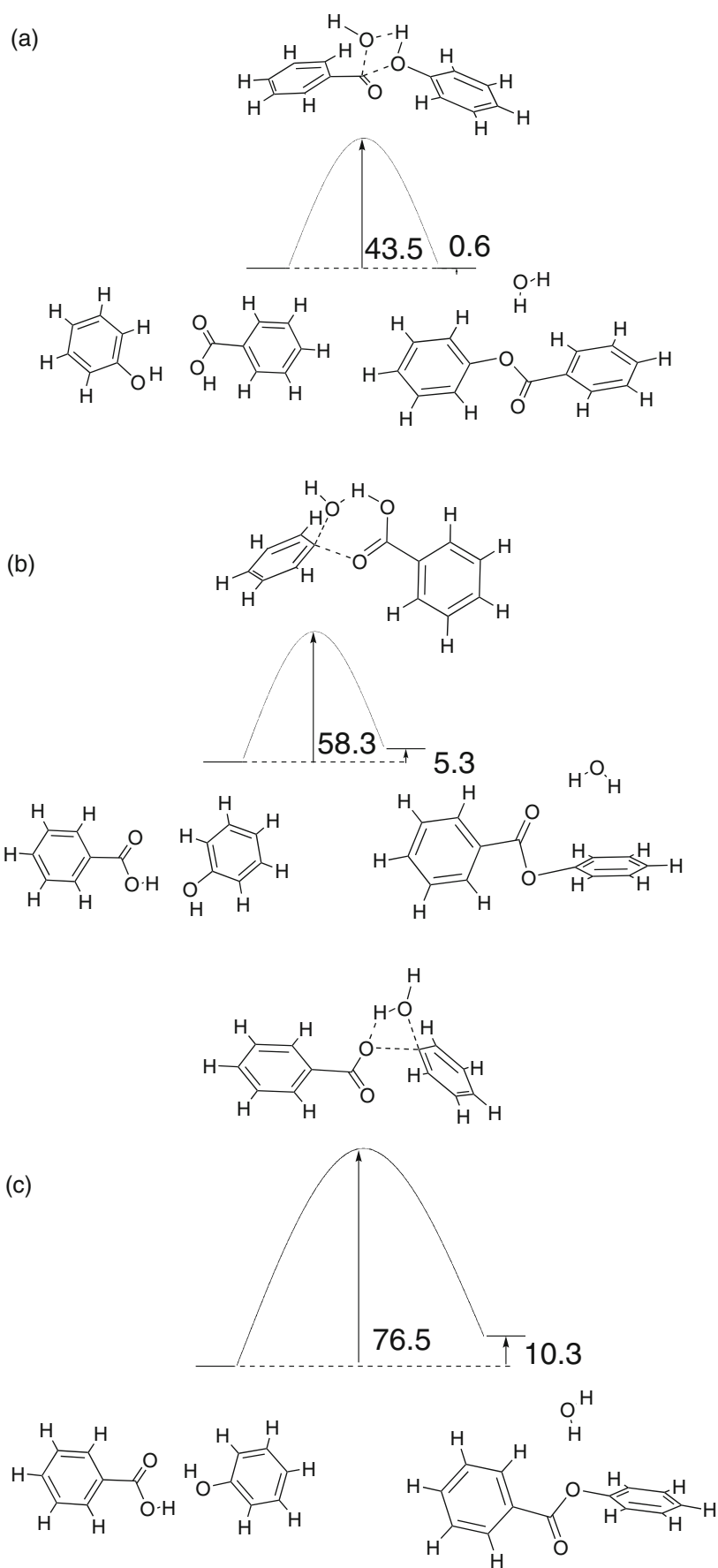
Bimolecular decarboxylation proceeds via the concerted pathway as shown in Fig. 2b. Its reaction mechanism can be regarded as an acid-promoted decarboxylation pathway. In this pathway, electrophilic attack of an acidic proton of one benzoic acid results in displacement of the CO₂ of the other benzoic acid. In the reactant complex, only one intermolecular hydrogen bond is formed in the benzoic acid dimer. One benzoic acid molecule plays a catalytic role in the proton transfer of the other benzoic acid. The activation barrier of this pathway is $36.8 \text{ kcal mol}^{-1}$.

It can be seen that, in these two bimolecular reaction mechanisms, the activation barrier of the first step of the stepwise coupling (i.e., dehydration) is lower than that of the second step of coupling reaction as well as that of the concerted bimolecular decarboxylation. Therefore, in the stepwise coupling mechanism, the further coupling reaction would be more difficult than anhydride formation due to the high activation barrier of the second step, although the decomposition of the two aryl radicals in the next step is easier. In addition, the formation of aryl anhydride is only slightly more favorable than the concerted decarboxylation mechanism. Thus, bimolecular dehydration and decarboxylation probably occur with comparable importance within the range of computational accuracy.

Esterification between carboxylic and hydroxyl groups

The esterification reaction mechanisms between the carboxylic groups and hydroxyl groups in brown coal using benzoic acid and phenol as model molecules are shown in Fig. 3. Three possible reaction mechanisms are proposed. In the first (Fig. 3a), phenyl benzoate is produced via elimination of the hydroxyl group of benzoic acid and the

Fig. 3 Energy diagrams of esterification reaction between benzoic acid and phenol (energy scale in kcal mol⁻¹)



proton of phenol with an activation barrier of 43.5 kcal mol⁻¹. The second and third reaction mechanisms assume direct water and ester formation through the combination of the proton of the carboxylic group of benzoic acid and the hydroxyl group of phenol, and C–O bond formation of the ester group. However, there are some differences in these two mechanisms. In the second pathway, C–O bond formation is due to combination of the carbonyl O and aryl C atoms, while in third, benzoic acid provides its proton to the hydroxyl group and contributes its oxygen atom from the hydroxyl group to form the C–O bond. The activation barrier of the second mechanism is 18.2 kcal mol⁻¹ lower than that of the third, mainly because the planar six-membered ring of the transition state in the second pathway is considerably less strained than the four-membered ring in the third pathway. The values of the activation barriers obtained from the esterification of benzoic acid with phenol indicated that the first reaction pathway is the dominant route.

Since brown coal is a solid macromolecular network, mass transport in coal is extremely restricted. Thus, the carboxylic groups existing in coal cannot change their positions easily like those in liquid phase. Most should be retained in their original isolated or hydrogen-bonded form before they perform pyrolysis reactions. Furthermore, the reactions occurring at carboxylic groups are not only controlled by their thermal barrier heights, but are also influenced by mass transport restrictions. The reaction of isolated single and hydrogen-bonded carboxylic groups should proceed to some extent independently.

The activation barriers of the predominant reaction pathways are summarized in Table 1 in the order of the height of activation barrier. From Table 1, it can be seen that the concerted bimolecular dehydration reaction, i.e., the anhydride formation reaction (Fig. 2a), has the lowest activation barrier and thus is the preferred reaction pathway if one carboxylic group can access the vicinity of another carboxylic group to form a hydrogen bond at low temperature in the first place. Further decomposition, which leads to aryl–aryl coupling from anhydride, is prohibited at this stage due to the high activation barrier (93.0 kcal mol⁻¹) that needs to be overcome. At the same time, a bimolecular concerted decarboxylation reaction may compete with the dehydration reaction as a result of comparable activation

barrier heights (36.8 kcal mol⁻¹). It should be noted that a single carboxylic group can be formed as the reaction product of bimolecular concerted decarboxylation, which can participate in the following thermal pyrolysis reaction as a unimolecular decarboxylation. However, due to the existence of restricted mass transport and the high activation barrier of unimolecular decarboxylation reactions (58.6 and 58.7 kcal mol⁻¹), most isolated single carboxylic groups should stay in their original configuration and participate only with difficulty in the thermal bimolecular decomposition reactions (dehydration and decarboxylation) at this stage. As the temperature increases, esterification between carboxylic groups and hydroxyl groups can occur if the carboxylic group is available to form a complex with the adjacent hydroxyl group, which is agreement with the results of Mae et al. [17]. The unimolecular decarboxylation of isolated single carboxylic groups becomes possible as the temperature increases due to its high activation barrier (58.7 and 58.6 kcal mol⁻¹). Furthermore, concerted and stepwise decarboxylation reactions can proceed simultaneously. The anhydrides formed from aromatic carboxylic group condensation at low temperature are precursors to aryl–aryl cross-links. However, bimolecular coupling should occur at higher temperature compared to other reactions mentioned above [17].

This preliminary investigation confirms that thermal decarboxylation of carboxylic groups, whether in bimolecular or unimolecular pathways, does not lead directly to cross-linking reactions in brown coals at low temperature, and hence CO₂ evolution from such reactions may not be related directly to cross-linking pathways, which is consistent with the results of Eskay et al. [4, 8]. Instead, bimolecular dehydration and esterification reactions with the release of H₂O can contribute to low temperature cross-links. The existence of hydrogen bonds and cross-linking reactions significantly affects the subsequent main pyrolysis reactions. These cross-linking reactions eliminated hydrogen-bonded carboxylic groups and therefore avoided the possibility that these groups undergo unimolecular decarboxylation like isolated single carboxylic groups at high temperature. In addition, it should be noted that decarboxylation occurred primarily by an acid-promoted bimolecular mechanism that does not lead to cross-linking but rather decomposition at low temperature. Only at

Table 1 Activation barriers of the possible predominant reaction pathways

Reaction	Mechanism	Activation barrier (kcal mol ⁻¹)
Bimolecular dehydration	Concerted	35.9
Bimolecular decarboxylation	Concerted	36.8
Esterification	Concerted	43.5
Unimolecular decarboxylation	Concerted	58.7
	Stepwise	58.6
Bimolecular coupling	Stepwise	93.0

high temperature, can the unimolecular decarboxylation of isolated single carboxylic groups and aryl–aryl coupling via the radical decarboxylation of dehydration products with the release of CO and CO₂ be achieved [18].

However, it is realized that simple model compounds may not be able to completely portray the complex chemistry and physical environment found in coal. For example, water pretreatment of brown coal at temperatures below the critical temperature has been shown to remove oxygen functional groups that are thought to be responsible for cross-linking reactions [19]. To complete our investigation into the role of water in thermal pyrolysis reactions, the possible thermal reaction mechanisms of carboxylic groups with the participation of water molecules are currently under investigation using quantum chemical calculations.

Conclusions

We have carried out a series of quantum chemical calculations for the possible reaction pathways between aromatic carboxylic groups themselves and with aromatic hydroxyl groups in a brown coal model. For hydrogen-bonded dimeric carboxylic groups, two competing reaction pathways, bimolecular dehydration and decarboxylation, were found and confirmed theoretically at lower temperatures. For hydrogen-bonded carboxylic and phenolic hydroxyl groups, the esterification reaction of the carboxylic group with the hydroxyl group can also proceed as the temperature rises due to its slightly higher activation barrier. For isolated single carboxylic groups, unimolecular decarboxylation in both concerted and stepwise pathways, and aryl–aryl coupling were all possible but only at higher reaction temperature. The results obtained also reveal that the main routes to the

low temperature cross-link are bimolecular dehydration and esterification reactions rather than bimolecular or unimolecular decarboxylation. Aryl–aryl coupling can contribute to cross-linking reactions only at high temperature via radical pyrolysis and recombination.

References

1. Domazetis G, Raoarun M, James BD (2006) *Energy Fuels* 20:1997–2007
2. Ozvatan S, Yurum Y (2002) *Energy Sources* 24:581–589
3. Li C-Z (2007) *Fuel* 86:1664–1683
4. Eskay TP, Britt PF, Buchanan AC (1996) *Energy Fuels* 10:1257–1261
5. Ibarra J, Moliner R, Gavilan MP (1991) *Fuel* 70:408–413
6. Joseph JT, Forrai TR (1992) *Fuel* 71:75–80
7. Manion JA, McMillen DF, Malhotra R (1996) *Energy Fuels* 10:776–788
8. Eskay TP, Britt PF, Buchanan AC (1997) *Energy Fuels* 11:1278–1287
9. Mae K, Maki T, Okutsu H, Miura K (2000) *Fuel* 79:417–425
10. Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) *Gaussian 03, Revision A.1.* Gaussian Inc, Pittsburgh PA
11. Ding L, Fang WH (2010) *J Org Chem* 75:1630–1636
12. Peng C, Ayala PY, Schlegel HB, Frisch MJ (1996) *J Comput Chem* 17:49–56
13. Peng C, Schlegel HB (1993) *Isr J Chem* 33:449–454
14. Gonzalez C, Schlegel HB (1990) *J Phys Chem* 94:5523–5527
15. Li J, Brill TB (2003) *J Phys Chem A* 107:2667–2673
16. Davidson D, Newman P (1952) *J Am Chem Soc* 74:1515–1516
17. Mae K, Maki T, Miura K (2002) *J Chem Eng Jpn* 35:778–785
18. Britt PF, Mungall WS, Buchanan AC (1998) *Energy Fuels* 12:660–661
19. Ross DS, Loo BH, Tse DS, Hirschon AS (1991) *Fuel* 70:289–295

Study of the betulin molecule in a water environment; ab initio and molecular simulation calculations

Miroslav Pospíšil · Petr Kovář · Robert Vácha ·
Michal Svoboda

Received: 1 September 2010 / Accepted: 22 March 2011 / Published online: 3 May 2011
© Springer-Verlag 2011

Abstract Ab initio and molecular simulation methods were used in calculations of the neutral individual betulin molecule, and molecular simulations were used to optimize the betulin molecule immersed in various amounts of water. Individual betulin was optimized in different force fields to find the one exhibiting best agreement with ab initio calculations obtained in the Gaussian03 program. Dihedral torsions of active groups of betulin were determined for both procedures, and related calculated structures were compared successfully. The selected force field was used for subsequent optimization of betulin in a water environment, and a conformational search was performed using quench molecular dynamics. The total energies of betulin and its interactions in water bulk were calculated, and the influence of water on betulin structure was investigated.

Keyword Betulin · Molecular simulation · Ab initio calculation · Hydrogen bond

Introduction

Betulin has long been investigated by research scientists due to its anticancer properties. Betulin and some of its

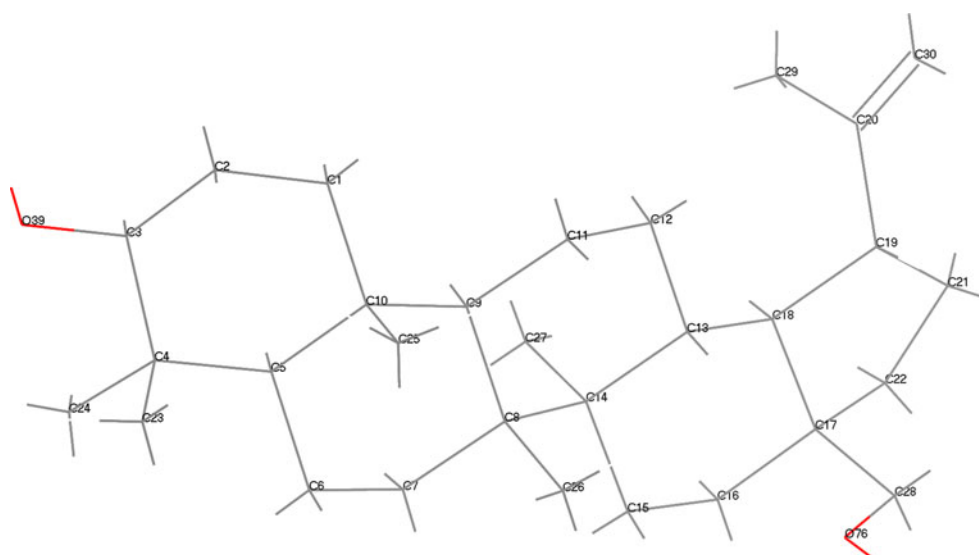
transformation products were first isolated from Maine white birch, *Betula papyrifera*, in the 1950s. Subsequently, betulinic acid was found to be uniquely effective against various types of human melanoma. This led to a search for other anticancer drugs from natural sources such as plants and marine organisms. As a result, new anticancer drugs like actidione (cycloheximide), carminomycin and bufalin were discovered and synthesized. Such synthesized anticancer substances followed known biosynthetic lead structures (2-haloethylamines, diazoketone groups). By contrast, marine invertebrates are a remarkable source of antineoplastic (cytostatic) substances and anticancer drugs with completely novel structural types. These newly discovered anticancer substances came in different modifications like cephalostatin, spongistatin, halistatin, dolastatin, bryostatin, etc. [1]. Betulin and its derivatives belong to the group of natural lupane compounds. The natural source of betulin is European white birch (*Betula pendula* Roth.), which is found mostly in Russia. Naturally occurring betulin can be isolated from many different sources such as trees {*Quercus suber* L [2], *Acacia mellifera* bark traditionally used in African medicine [3], *Celtis philippinensis* Blanco (Indonesia) [4], aerial parts of *Asteracantha longifolia* Nees (India) [5], aerial parts of *Coccoloba acrostichoides* Cham (Brazil) [6], *Betula papyrifera* bark} by vacuum and atmospheric sublimation, with the optimal conditions for extraction being 3.3 kPa and isothermal heating at 200°C [7].

Synthesis of betulin and its derivatives is in great demand due to its biological activity. Betulin and its esters can inhibit growth of cancer cells and prevent tumor development, suppress HSV-1 and HSV-2 replication, cause 50% inhibition of HIV-1 replication as an active anti-HIV agent, and exhibit some antitubercular activity [8–11]. Betulin derivatives and isomers often have better desirable properties for the pharmaceutical industry than betulin

M. Pospíšil (✉) · P. Kovář
Faculty of Mathematics and Physics,
Charles University in Prague,
Ke Karlovu 3,
12116 Prague 2, Czech Republic
e-mail: pospasil@karlov.mff.cuni.cz

M. Pospíšil · R. Vácha · M. Svoboda
I.Q.A.,
Jindřicha Plachty 16,
15000 Prague 5, Czech Republic

Fig. 1 Betulin molecule optimized in the Compass force field [44]. Red Oxygen, gray carbon, white with outline hydrogen



itself. For example, they can be applied as: (1) anti HIV agents 3,28-di-O-(dimethylsuccinyl)-betulin isomers [12] and 3-O-glutaryl-dihydrobetulin with related monoacyl derivatives [13], (2) antiviral active components against herpes simplex virus [14, 15], (3) anti-inflammatory and antiulcer properties of betulin bis-hemiphthalate [16], (4) anticancer drug candidates [17–20], (5) agents with inhibitory effect against the intracellular bacterium *Chlamydia pneumoniae* [21] etc.

The anti-inflammatory, antitumor, immunomodulating and anti-HIV activities of betulin and its derivatives have been investigated at the macro scale and measured by different experimental techniques, including NMR, FTIR, UV VIS, X-ray, and reversed-phase thin-layer chromatography [22–33]. By testing different kinds of derivatives it has been shown that the properties observed at a macro scale are influenced strongly at the nano scale level by the C-3 as well as the C-28 atomic position of betulin. The molecular and crystal structure of betulin and its derivatives have been described in various solvates, revealing the important influence of non-bond interactions related to OH groups [34, 35]. The C-3 and C-28 atoms bear OH groups whose position, mutual distance and orientation with

respect to the ring part of the betulin molecule can influence hydrogen bonding and the interactions of betulin with active sites of surrounding species and thus its biological activity [36]. For these reasons, and to gain a better understanding of the structure–property relationships of this important molecule, it is appropriate to investigate the structure of betulin at the nanoscale level using computer calculation techniques. Recent progress in computer molecular simulations have opened up the possibility of screening different kinds of betulin derivatives to search for chemically accessible substances to obtain the best candidate by combining high activity and chemical accessibility [37]. For example, the influence of betulin with respect to the anti-phospholipase A₂ activity was investigated by molecular modeling methods. It was shown, and confirmed experimentally, that betulin is responsible for the observed anti-phospholipase A₂ activity [38]. Interactions of betulin and betulinic acid were investigated in cyclodextrine complexes by experimental methods and quantum chemical calculations [32, 39, 40].

In the present work, the betulin molecule was calculated and optimized by ab initio calculation in Gaussian03 [41],

Table 1 Examples of selected optimized models of the betulin molecule with minimum energy in different force fields

Force field/energy (kcal/mol)	Total	Bonds	Angles	Torsions	van der Waals	Coulomb
Compass	-40.6	7.4	25.9	-76.4	22.5	-10.7
CFF91	-23.6	11.1	27.4	-83.9	41.9	-11.3
UFF	131.7	31.0	45.1	9.0	56.7	-10.1
CVFF	136.9	293.7	26.7	12.1	78.1	-9.7
Dreiding2.21	154.6	31.4	35.7	8.5	85.8	-6.9

and these parameters were used to perform molecular dynamics (MD) calculations in the Gromacs simulation package [42]. These results were subsequently compared with molecular simulation results from the Cerius²/Material Studio modeling environment [43] under different force fields and various initial conditions or conformations. The optimized model of the betulin molecule in vacuum was subsequently inserted into a water environment. No previous study has examined water-influenced structural changes in betulin. The aims of the study were: (1) to find the best description of betulin structure by comparing the results from MD simulations packages under various forcefield parameters and results based on ab initio calculations; (2) to determine the influence of water under various concentrations, and determine a number of dynamics steps to save computer time to obtain relevant results; (3) to determine the structural changes to betulin caused by water bulk with real density. These results and procedures can be used for further investigation of the behavior of betulin and its derivatives in order to gain a better understanding of their biological activity.

Molecular calculations

The initial model sets of the betulin molecule were created in 3D sketcher module [43] with different conformations. The individual molecules of betulin were optimized and their energy minimized in different force fields to select the conformation most suitable for subsequent calculations with respect to ab initio calculation results. The optimization of initial models was done in Compass [44], CFF91 [45, 46], UFF [47–49], CVFF-generalized valence [50] and

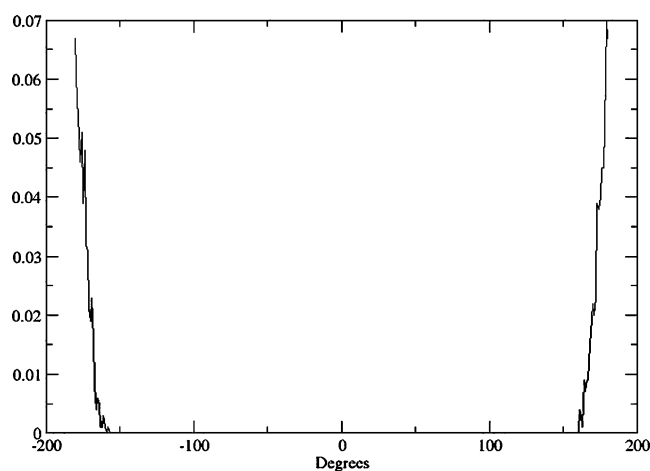


Fig. 3 Dihedral distribution of CH₂ in C₃H₅ groups based on ab initio parameters of the betulin molecule

Dreiding 2.21 [51] force fields (Fig. 1). Charges were calculated by the charge equilibration method (Qeq) [52]. Different optimization methods ranging from the less precise steepest descent, conjugate gradient to the most sophisticated Newton Raphson methods were used to optimize all sets of initial models. The resultant models were then compared with the results from ab initio calculations, and the model with the best agreement was selected. This model was inserted into a bulk with different numbers of water molecules (67, 134 and 480) to generate new initial model sets. These initial models were again optimized without any restriction in Compass force fields. The following conditions were used after the first energy minimization: the quenched dynamics started in an NVT ensemble for T=300 K, and a dynamic time step of

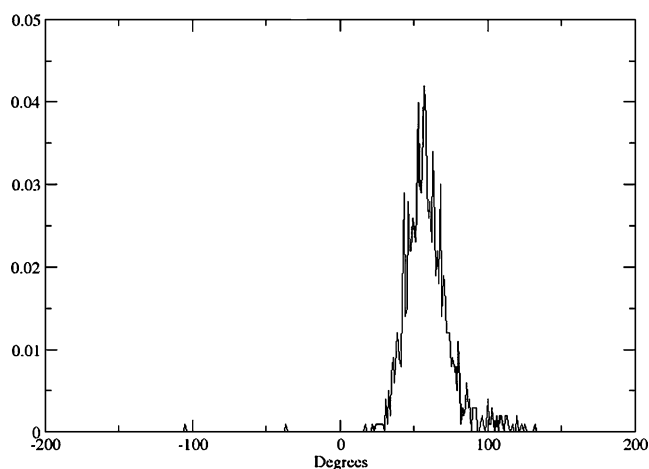


Fig. 2 Dihedral distribution of C₃H₅ groups based on ab initio parameters of the betulin molecule

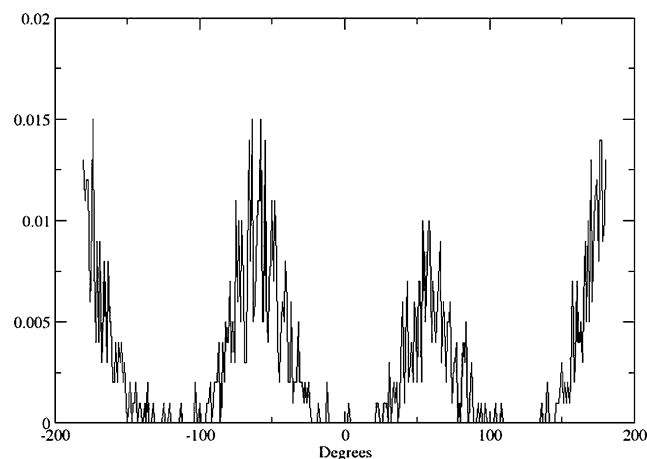


Fig. 4 Dihedral distribution of CH₃ in C₃H₅ groups based on ab initio parameters of the betulin molecule

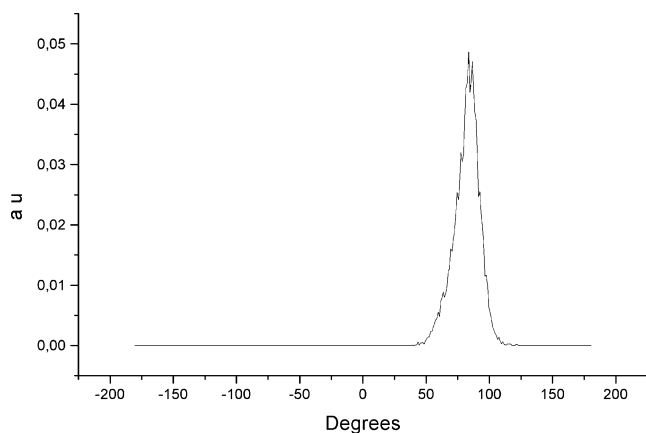


Fig. 5 Dihedral distribution of C_3H_5 groups calculated by molecular dynamics (MD)

0.0001 ps; the temperature was kept constant using a Berendsen thermostat [53].

After a dynamics run of 5,000 ps, and after reaching an equilibrium state (approximately 2,000 ps), the structures at times of 3,000 ps, 4,000 ps and 5,000 ps were selected and optimized to obtain the minimum energy. The total bond energy was represented by analytical potentials, which depend on the type of forcefield, and energy consisted of bond, angle, torsion, inversion and cross terms [43, 54]. The total non-bond energy involved Coulomb and Van der Waals energies [43, 54]. The cutoff (0.950 nm) and spline (0.100 nm) were used to calculate the Coulombic energy. Van der Waals energy was computed using the Lennard-Jones functional form [55].

The betulin molecule was optimized by the Hartree-Fock method using 6-31 G* basis in Gaussian03 [41]. The electrostatic field around the molecule was calculated according to the Merz-Singh-Kollman scheme, from which

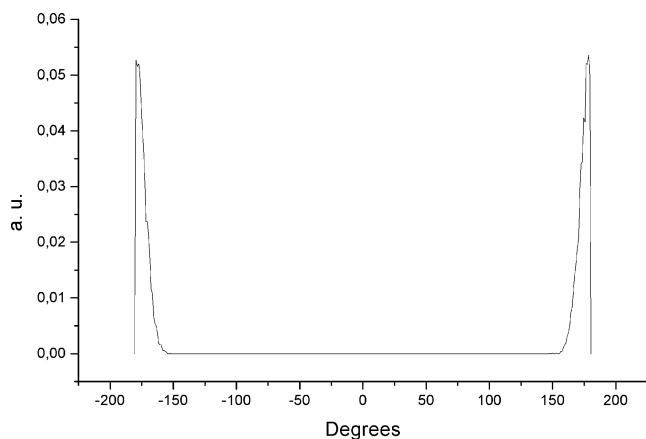


Fig. 6 Dihedral distribution of CH_2 in C_3H_5 groups calculated by MD

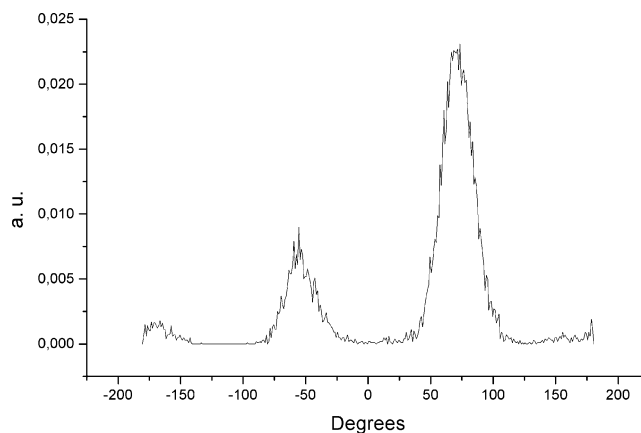


Fig. 7 Dihedral distribution of CH_3 in C_3H_5 groups calculated by MD

the atomic point charges were calculated using the RESP method [56, 57]. Force field parameters were obtained and based on GAFF (generalized amber force field) [58]. This force field was used in the Gromacs simulation package for dynamic calculations [42].

Results and discussion

Calculation techniques

The results of optimization of initial models with various conformations of betulin in different force fields are presented in Table 1. A comparison between the results of ab initio methods and classical calculations are presented in Figs. 2, 3, 4, 5, 6, 7, 8, and 9. The figures show dihedral torsions of CH_3 , CH_2 , C_3H_5 group and OH groups. For simplicity, the graphs of dihedral torsions from ab initio

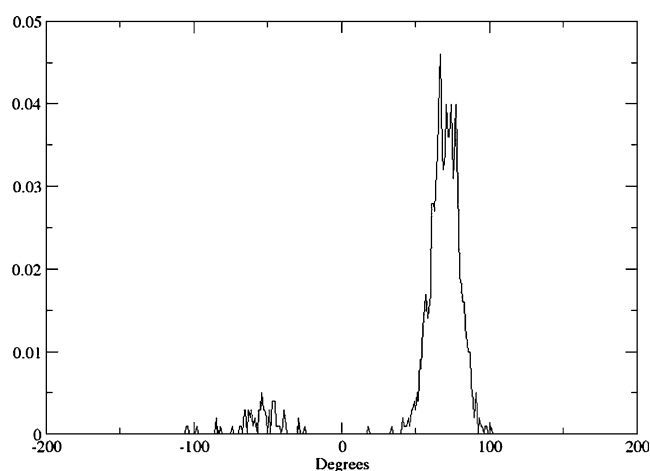


Fig. 8 Dihedral distribution of HCOH groups near C_3H_5 groups based on ab initio parameters of the betulin molecule

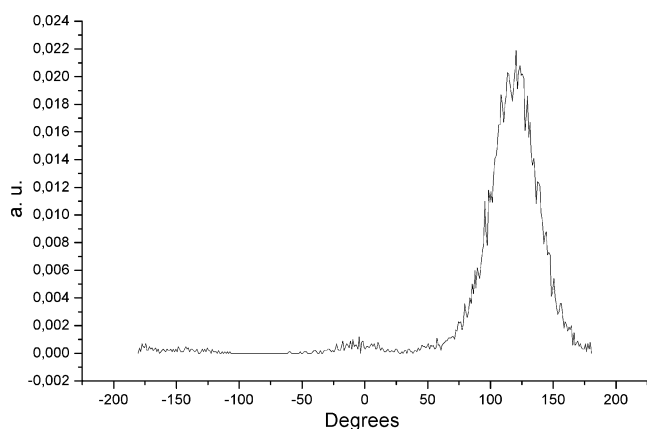


Fig. 9 Dihedral distribution of HOCH groups near C_3H_5 groups calculated by MD

calculations are framed. If we compare Figs. 2 and 5, 3 and 6, 4 and 7, 8 and 9, we can see a good agreement between the positions of the peaks, indicating the similar behavior of betulin in both calculation procedures. The intensity differences between Figs. 4 and 7 can be explained by the presence of local energy barriers that were impossible to overcome during classical MD.

On the basis of agreement between *ab initio* and molecular mechanics results, and the results presented in Table 1 (on the basis of total energy values), the Compass

force field was selected for subsequent optimization, energy calculation and description of betulin in a water environment containing various amounts of water molecules. The most important parts of the betulin molecule for the investigation in water are the two hydroxyl groups and the C_3H_5 group connected to the ring part of betulin composed of cyclopentane and cyclohexanes (see Fig. 1). Their mutual positions and torsions with respect to the ring part of betulin influence the possible connectivity of betulin and its biological activity. For this reason, attention was focused on the dihedral torsion of these active groups of betulin and their possible positions and interactions with respect to their surroundings.

Study of betulin in a water environment

The betulin molecule optimized in the Compass force field was immersed into a bulk with various amounts of water. After the first optimization step, classical MD were performed for 5,000 ps under the conditions stated above. The progress of MD was saved into a trajectory file every 1 ps. After the first part of the dynamics run, when the system reached the equilibrium state, the second collecting phase started. The system converged into the equilibrium state during the first 2,000 ps of the dynamics run. Figures 10, 11, and 12 show systems with 67, 134 and 480 water molecules, respectively, at a time of 5,000 ps, with dashed lines representing hydrogen bonds. As we can

Fig. 10 Betulin molecule immersed in 67 molecules of H_2O after a 5,000-ps MD run. Dashed lines Hydrogen bonds, grey C, red O, white H

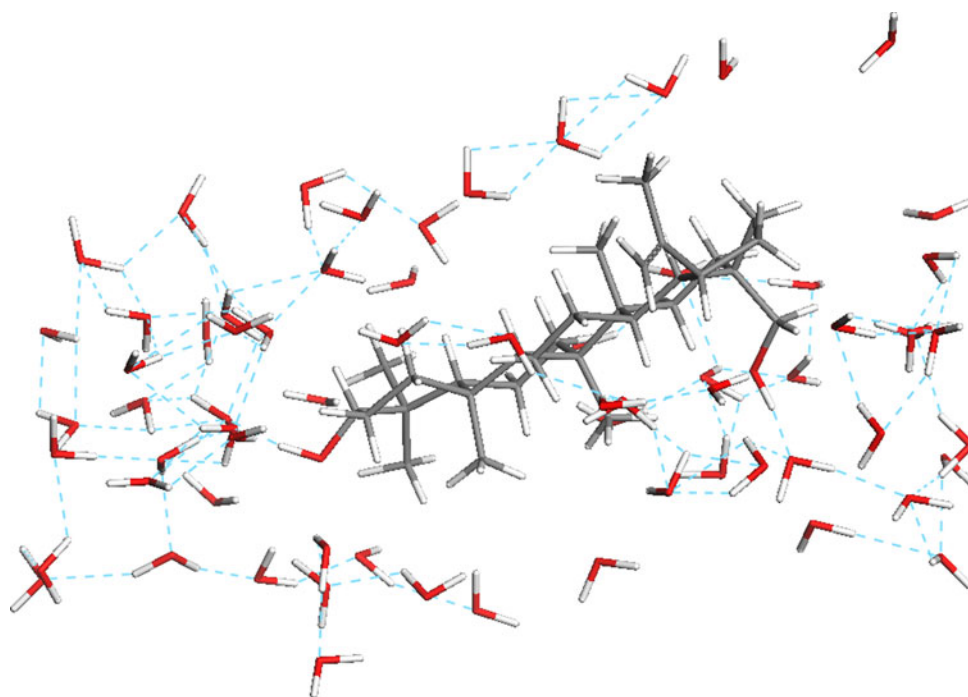
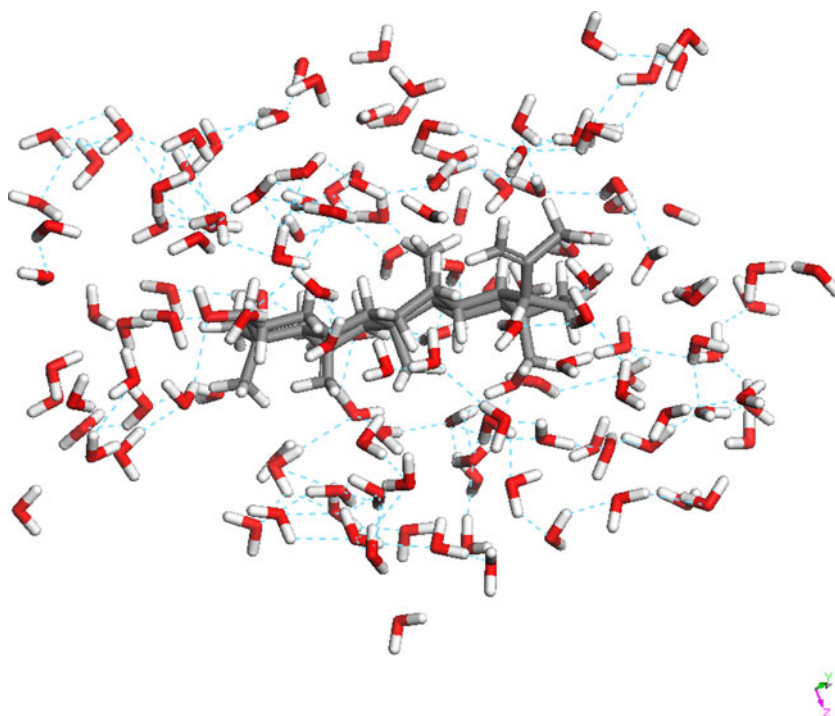


Fig. 11 Betulin molecule immersed in 134 molecules of H₂O after a 5,000-ps MD run. *Dashed lines* Hydrogen bonds, *grey* C, *red* O, *white* H



see, the active OH groups of betulin are surrounded by water molecules and hydrogen bonds are created. The other part of betulin is hydrophobic, with no hydrogen bonds (best seen in Fig. 12 due to the lower number of water molecules presented in the model for clarity).

Table 2 shows the influence of water on the structure of the betulin molecule. We compared dihedral torsion angles for betulin rings and for the C₃H₅ group (the last four rows of Table 2) for various amounts of water (0, 67, 134 and 480 water molecules). The last two columns of Table 2 represent the average values of torsion dihedral angles, and the absolute value of their maximal deviations. One can see that the most rigid parts of the molecule are C14–C15–C16–C17–C18–C13 and C17–C18–C19–C21–C22 rings. Maximal deviations for these rings are lowest in comparison to other rings in the molecule, with values of 9° and 11°, respectively. On the other hand, a part of betulin molecule that exhibits the biggest changes in dihedral torsion values is the C8–C9–C11–C12–C13–C14 ring. The changes in dihedral torsions are attributed to changes in the position of the C9 atom (see Fig. 13). This figure shows a side view of the betulin molecule optimized in vacuum (a black sphere represents the C9 atom). One can see that, in this case, the C8–C9–C11–C12–C13–C14 ring adopts a chair conformation. The change in position of the C9 atom due to interactions with water molecules leads to a saddle conformation of the C8–C9–C11–C12–C13–C14 ring and to the observed differences in dihedral torsion values.

The C5–C6–C7–C8–C9–C10 and C1–C2–C3–C4–C5–C10 rings exhibit comparable changes in dihedral torsion values (average maximal deviation is 16° and 17°, respectively), which are attributed to the conformational changes in the C8–C9–C11–C12–C13–C14 ring, and also probably to hydrophobic interactions of the betulin mole-

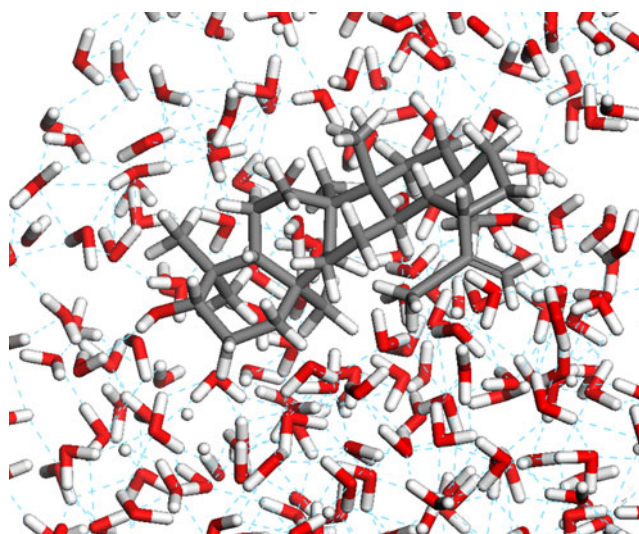


Fig. 12 Betulin molecule immersed in 480 molecules of H₂O after a 5,000-ps MD run. *Dashed lines* Hydrogen bonds, *grey* C, *red* O, *white* H. Water molecules above the betulin molecule are partially removed for clarity

Table 2 Comparison of dihedral torsion values of betulin for various concentrations of water

Torsion angle	Vacuum [°]	67 H ₂ O [°]	134 H ₂ O [°]	480 H ₂ O [°]	Average [°]	Maximum deviation [°]
C1-C2-C3-C4	56	40	69	42	52	17
C2-C3-C4-C5	-53	-23	-68	-55	-50	27
C3-C4-C5-C10	52	29	51	72	51	22
C4-C5-C10-C1	-50	-47	-47	-60	-51	9
C5-C10-C1-C2	51	59	48	37	49	11
C10-C1-C2-C3	-56	-58	-55	-34	-51	17
C5-C6-C7-C8	58	63	60	67	62	5
C6-C7-C8-C9	-46	-57	-52	-76	-58	18
C7-C8-C9-C10	44	48	46	70	52	18
C8-C9-C10-C5	-51	-48	-57	-39	-49	9
C9-C10-C5-C6	58	54	66	22	50	28
C10-C5-C6-C7	-63	-61	-61	-39	-56	17
C8-C9-C11-C12	57	61	57	-52	31	83
C9-C11-C12-C13	-55	-59	-66	-11	-48	37
C11-C12-C13-C14	54	55	69	62	60	9
C12-C13-C14-C8	-56	-61	-63	-51	-58	7
C13-C14-C8-C9	59	61	59	-8	43	51
C14-C8-C9-C11	-59	-58	-54	59	-28	87
C14-C15-C16-C17	52	55	53	49	52	3
C15-C16-C17-C18	-51	-50	-64	-56	-55	9
C16-C17-C18-C13	56	44	61	61	55	11
C17-C18-C13-C14	-58	-41	-61	-60	-55	14
C18-C13-C14-C15	52	44	55	61	53	9
C13-C14-C15-C16	-51	-53	-45	-53	-50	6
C17-C18-C19-C21	34	22	21	18	24	10
C18-C19-C21-C22	-5	13	10	22	10	15
C19-C21-C22-C17	-25	-46	-37	-50	-40	15
C21-C22-C17-C18	45	59	51	64	55	10
C22-C17-C18-C19	-49	-49	-46	-50	-48	3
C18-C19-C20-C29	-43	-56	-93	-51	-61	32
C29-C20-C19-C21	74	57	21	173	81	92
C18-C19-C20-C30	141	136	84	132	123	40
C30-C20-C19-C21	-102	-110	-162	-5	-95	90

cule with its surroundings. As one can also see from Table 2, the orientation of the C₃H₅ group with respect to C17–C18–C19–C21–C22 ring is very sensitive to the

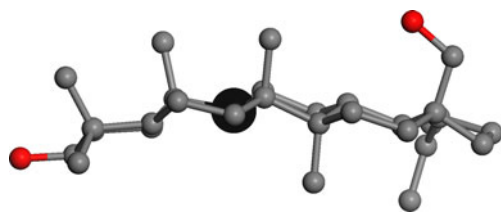


Fig. 13 Side view of the betulin molecule optimized in vacuum. Black sphere C9 atom, red spheres oxygen atoms

presence of water, and it can rotate freely around the C20–C19 axis in the range from ca. -70° to $+70^\circ$ with respect to the C₃H₅ position of betulin optimized in vacuum.

It is worth noting that changes in the geometry of the betulin molecule caused by the interactions with its

Table 3 Interactions of the betulin molecule in a water environment

Energy	0 H ₂ O	67 H ₂ O	134 H ₂ O	480 H ₂ O
E_b	-40	33	112	265
E_i	0	-250	-680	-849

surroundings lead to variability in the distances between the C3 and C28 and O39 and O76 atoms. In general, the presence of water leads to a contraction of the C3–C28 distance from 10.4 Å for 0 H₂O to 8.9 Å for 480 H₂O. On the other hand, the distance between the O39 and O76 atoms ranges randomly between 10.4 Å and 11.5 Å. This can be attributed to the fact that the OH group linked to the C28 atom can rotate freely around C18–C17 bond.

The interactions of the betulin molecule are described in Table 3. It shows the total energy, E_b , for the isolated betulin molecule, and the sum of non-bonded interactions, E_i , of betulin with water. The optimized betulin molecule without a water environment exhibits the best value of total energy, which increases rapidly to positive values of E_b with increasing concentrations of water, and can better illustrate the consequence of conformational changes of the molecule described by changes in dihedral torsions. Based on the results in Table 3, we can conclude that the influence of water leads only to deformations of the betulin molecule and not to a stable conformation corresponding to a local minimum with a negative value of total energy E_b . The increase in energy, E_b , can be attributed to the fact that increasing amounts of water molecules lead to stronger mutual interactions between the betulin molecule and water, which leads to higher deformations of betulin.

The total energy and its components for the selected models at times of 3,000 ps, 4,000 ps and 5,000 ps are presented in Table 4. The total relative energy decreases with increasing amounts of water molecules in the system. The model with 480 water molecules represents betulin immersed in a bulk of water with surroundings close to the real density of water (1,000 kg/m³). The water bulk has a spherical shape with a radius, $r=1.5$ nm. The results show that it is possible to calculate shorter MD trajectories due to negligible changes in energy values for dynamics runs of

3,000–5,000 ps. This is important for molecular simulation calculations of betulin derivatives or isomers because of significant computation time savings.

Based on the results presented, one can see that the influence of the water environment is very important and cannot be neglected in the description of betulin and investigation of its properties. Our results show that the presence of water can be important for the complex behavior of betulin during theoretical studies investigating the solubility of betulin.

Summary

The betulin molecule is important in the pharmaceutical industry due to its unique properties in many therapeutic applications. Molecular simulations can bring new important information on betulin behavior in water environments. A minimum set of necessary parameters was established for different amounts of water molecules. The results confirm the influence of two hydroxyl groups for the bonding properties of betulin. Torsions related to C3 and C28 were determined on models calculated by molecular simulations and ab-initio-based methods, and a very good agreement was obtained in Compass force field [44]. From the results based on MD simulations, we can conclude that water has a significant influence on the geometry of the betulin molecule. The main changes can be observed in the C14–C15–C16–C17–C18–C13 ring, where the chair conformation and the saddle conformation are possible. These changes and interactions with water molecules lead to the distortion of betulin molecule and this is accompanied by the contraction of C3–C28 distance.

Nevertheless, many limitations, such the amount of water molecules, still exist, limiting a more precise

Table 4 Comparison of betulin under different concentrations of water molecules in the Compass force field [44] after 3,000 ps, 4,000 ps and 5,000 ps of molecular dynamics (MD)

Force field, amount of water, time / energy [kcal/mol]	Total	Bond	Angle	Torsion	Cross	van der Waals	Coulomb
Compass, 67×H ₂ O, 3,000 ps	-4,242.3	85.9	132.2	-70.7	-22.7	208.7	-4,575.7
Compass, 67×H ₂ O, 4,000 ps	-4,252.0	86.8	130.7	-70.3	-21.9	205.1	-4,582.3
Compass, 67×H ₂ O, 5,000 ps	-4,253.6	87.0	130.3	-70.3	-21.9	207.1	-4,585.7
Compass, 134×H ₂ O, 3,000 ps	-9,878.3	219.7	252.6	-61.2	-34.3	194.5	-10,449.6
Compass, 134×H ₂ O, 4,000 ps	-9,889.2	219.2	248.7	-61.9	-34.8	190.1	-10,450.6
Compass, 134×H ₂ O, 5,000 ps	-9,867.1	219.0	244.1	-62.1	-34.9	185.6	-10,418.9
Compass, 480×H ₂ O, 3,000 ps	-48,838.1	1,714.5	1,002.3	-60.5	-175.5	2,055.0	-53,374.0
Compass, 480×H ₂ O, 4,000 ps	-48,828.8	1,647.9	1,013.0	-59.7	-167.6	2,103.1	-53,365.6
Compass, 480×H ₂ O, 5,000 ps	-48,662.6	1,730.8	999.2	-61.0	-171.5	2,155.2	-53,315.3

description and calculation of betulin and other molecules. With increased computing resources, we hope to obtain more sophisticated results and a deeper understanding of the investigated properties and behavior of betulin and similar molecules.

Acknowledgments The work was performed thanks to the financial support of the Ministry of Industry and Trade of the Czech Republic project number: FT-TA/027 and maintenance technical support from the Ministry of Education, Youth and Sports of the Czech Republic (Project MSM 0021620835).

References

- Pettit GR (1996) *J Nat Prod* 59:812–821
- Castola V, Bighelli A, Rezzi S, Melloni G, Gladiali S, Desjoberg JM, Casanova J (2002) *Ind Crops Prod* 15:15–22
- Mutai C, Abatis D, Vagias C, Moreau D, Roussakis C, Roussis V (2004) *Phytochemistry* 65:1159–1164
- Hwang BY, Chai HB, Kardono LBS, Riswan S, Farnsworth NR, Cordell GA, Pezzuto JM, Kinghorn AD (2003) *Phytochemistry* 62:197–201
- Misra TN, Singh RS, Pandey HS, Singh BK, Pandey RP (2001) *Fitoterapia* 72:194–196
- Cota BB, de Oliveira AB, de Souza-Filho JD, Braga FC (2003) *Fitoterapia* 74:729–731
- Guidoin MF, Yang J, Pichette A, Roy C (2003) *Thermochim Acta* 398:153–166
- Tolstikova TG, Sorokina IV, Tolstikov GA, Tolstikov AG, Flekhter OB (2006) *Russian J Bioorg Chem* 32:37–49
- Rzeski W, Stepulak A, Szymański M, Juszcak M, Grabarska A, Siffringer M, Kaczor J, Kandefér-Szerszeń M (2009) *Basic Clin Pharmacol Toxicol* 105:425–432
- Kommerer H, Kaluderović GN, Kalbitz J, Pachke R (2011) *Invest New Drugs* 29:266–272. doi:10.1007/s10637-009-9358-x
- Sami A, Taru M, Salme K, Jari YK (2006) *Eur J Pharm Science* 29:1–13
- Kashiwada Y, Chiyo J, Ikeshiro Y, Nagao T, Okabe H, Cosentino LM, Fowke K, Lee KH (2001) *Bioorg Med Chem Lett* 11:183–185
- Kashiwada Y, Sekiya M, Ikeshiro Y, Fujioka T, Kilgore NR, Wild CT, Allaway GP, Lee KH (2004) *Bioorg Med Chem Lett* 14:5851–5853
- Pavlova NI, Savinova OV, Nikolaeva SN, Boreko EI, Flekhter OB (2003) *Fitoterapia* 74:489–492
- Baltina LA, Flekhter OB, Nigmatullina LR, Boreko EI, Pavlova NI, Nikolaeva SN, Savinova OV, Tolstikov GA (2003) *Bioorg Med Chem Lett* 13:3549–3552
- Karachurina LT, Sapozhnikova TA, Zarudii FS, Flekhter OB, Galin FZ (2002) *Pharm Chem J* 36:432–433
- Šarek J, Klinot J, Džubák P, Klinotová E, Nosková V, Křeček V, Kořínková G, Thomson JO, Janošková A, Wang S, Parsons S, Fischer PM, Zhelev NZ, Hajdúch M (2003) *J Med Chem* 46:5402–5415
- Kvasnica M, Šarek J, Klinotová E, Džubák P, Hajdúch M (2005) *Bioorg Med Chem* 13:3447–3454
- Mar AA, Koohang A, Majewski ND, Szotek EL, Eizhamer DA, Flavin MT, Xu ZQ (2009) *Chin Chem Lett* 20:1141–1144
- Kommerer H, Kaluderović GN, Kalbitz J, Pachke R (2010) *Arch Pharm Chem Life Sci* 8:449–457
- Salin O, Alakurtti S, Pohjala L, Siiskonen A, Maass V, Maass M, Yli-Kauhaluoma J, Vuorela P (2010) *Biochem Pharmacol* 80:1141–1151
- Uzenkova NV, Petrenko NI, Shakirov MM, Shul'ts EE, Tolstikov GA (2005) *Chem Nat Compd* 41:692–700
- Tolmacheva IA, Shelepen'kina LN, Vikharev YB, Anikina LV, Grishko VV, Tolstikov AG (2005) *Chem Nat Compd* 41:701–705
- Flekhter OB, Giniyatullina GV, Galin FZ, Baschenko NZ, Makara NS, Zarudii FS, Boreko EI, Savinova OV, Pavlova NI, Starikova ZA, Tolstikov GA (2005) *Chem Nat Compd* 41:706–709
- Flekhter OB, Medvedeva NI, Karachurina LT, Baltina LA, Galin FZ, Zarudii FS, Tolstikov GA (2005) *Pharm Chem J* 39:401–404
- Flekhter OB, Karachurina LT, Nigmatullina LR, Sapozhnikova TA, Baltina LA, Zarudii FS, Galin FZ, Spirikhin LV, Tolstikov GA, Plyasunova OA, Pokrovskii AG (2002) *Russ J Bioorg Chem* 28:494–500
- Symon AV, Veselova NN, Kaplun AP, Vlasenkova NK, Fedorova GA, Lyutik AI, Gerasimova GK, Shvets VI (2005) *Russ J Bioorg Chem* 31:286–291
- Flekhter OB, Medvedeva NI, Karachurina LT, Baltina LA, Zarudii FS, Galin FZ, Tolstikov GA (2002) *Pharm Chem J* 36:488–491
- Korovin AV, Tkachev AV (2001) *Russ Chem Bull Int Ed* 50:304–310
- Sun IC, Shen JK, Wang HK, Cosentino LM, Lee KH (1998) *Bioorg Med Chem Lett* 8:1267–1272
- Sun IC, Wang HK, Kashiwada Y, Shen JK, Cosentino LM, Chen CH, Yang LM, Lee KH (1998) *J Med Chem* 41:4648–4657
- Fälamaş A, Pinzaru SC, Dehelean CA, Peev CI, Soica C (2011) *J Raman Spectrosc* 42:97–107
- Achem-Achremowicz J, Kępczyńska E, Zylewski M, Janeczko Z (2009) *Biomed Chromatogr* 24:261–267
- Trishin YG, Chernyavskii GG, Shafeeva MV, Nelyubina YV (2010) *Russ J Org Chem* 46:1490–1492
- Drebushchak TN, Mikhailenko MA, Brezgunova ME, Shakhtshneider TP, Kuznetsova SA (2010) *J Struct Chem* 51:798–801
- Hiroya K, Takahashi T, Miura N, Naganuma A, Sakamoto T (2002) *Bioorg Med Chem* 10:3229–3236
- Pogrebnyak AV, Vasilenko YuK, Oganessian ÉT, Glushko AA, Suzdalev KF, Pogrebnyak LV (2002) *Pharm Chem J* 36:535–537
- Bernard P, Scior T, Didier B, Hibert M, Berthon JY (2001) *Phytochemistry* 58:865–874
- Soica CM, Peev CI, Ciurlea S, Ambrus R, Dehelean C (2010) *Farmacia* 58:611–619
- Falamaş A, Pinzaru SC, Chiş V, Dehelean C (2010) *J Mol Struct*. doi:10.1016/j.molstruc.2010.11.044
- Frisch MJ et al (2004) Gaussian 03, Revision C.02. Gaussian Inc, Wallingford CT
- Lindahl E, Hess B, van der Spoel D (2001) *J Mol Model* 7:306–317
- Accelrys Software Inc, Cerius Modeling Environment, Release 4.5/Material Studio Modeling Environment, Release 4.5 (2003) Accelrys Software Inc, San Diego
- Sun H (1998) *J Phys Chem B* 102:7338–7364
- Maple JR, Hwang MJ, Stockfisch TP, Dinur U, Waldman M, Ewig CS, Hagler AT (1994) *J Comput Chem* 15:162–182
- Hwang MJ, Stockfisch TP, Hagler AT (1994) *J Am Chem Soc* 116:2515–2525
- Rappé AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM (1992) *J Am Chem Soc* 114:10024–10035
- Casewit CJ, Colwell KS, Rappé AK (1992) *J Am Chem Soc* 114:10035–10046
- Rappé AK, Colwell KS, Casewit CJ (1993) *Inorg Chem* 32:3438–3450
- Dauber-Osguthorpe P, Roberts VA, Osguthorpe DJ, Wolff J, Genest M, Hagler AT (1988) *Proteins Struct Funct Genet* 4:31–47

51. Mayo SL, Olafson BD, Goddard WA III (1990) *J Phys Chem* 94:8897–8909
52. Rappe AK, Goddard WA III (1991) *J Phys Chem* 95:3358–3363
53. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) *J Chem Phys* 81:3684–3690
54. Comba P, Hambley TW (1995) *Molecular modelling of inorganic compounds*. VCH, Weinheim
55. Lennard-Jones JE (1925) *Proc R Soc London Ser A* 109:584
56. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) *J Phys Chem* 97:10269–10280
57. Cornell WD, Cieplak P, Bayly CI, Kollman PA (1993) *J Am Chem Soc* 115:9620–9631
58. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) *J Comput Chem* 25:1157–1174

Molecular dynamics simulation and density functional theory studies on the active pocket for the binding of paclitaxel to tubulin

Sichuan Xu · Shaoming Chi · Yi Jin · Qiang Shi ·
Maofa Ge · Shu Wang · Xingkang Zhang

Received: 7 January 2011 / Accepted: 5 April 2011 / Published online: 3 May 2011
© Springer-Verlag 2011

Abstract Paclitaxel (PTX) is used to treat various cancers, but it also causes serious side effects and resistance. To better design similar compounds with less toxicity and more activity against drug-resistant tumors, it is important to clearly understand the PTX-binding pocket formed by the key residues of active sites on β -tubulin. Using a docking method, molecular dynamics (MD) simulation and density functional theory (DFT), we identified some residues (such as Arg278, Asp26, Asp226, Glu22, Glu27, His229, Arg369, Lys218, Ser277 and Thr276) on β -tubulin that are the active sites responsible for interaction with PTX. Another two residues, Leu371 and Gly279, also likely serve as active sites. Most of these sites contact with the “southern hemisphere” of PTX; only one key residue interacts with the “northern hemisphere” of PTX. These key residues can be divided into four groups, which serve as active compositions in the formation of an active pocket for PTX binding to β -tubulin. This active binding pocket enables a very strong interaction (the strength is predicted to be in the range of -327.8 to -365.7 kJ mol⁻¹) between β -tubulin and PTX, with various orientated conformations. This strong interaction means that PTX possesses a high level of activity against cancer cells, a result that is in good agreement with the clinical mechanism of PTX. The

described PTX pocket and key active residues will be applied to probe the mechanism of tumor cells resistant to PTX, and to design novel analogs with superior properties.

Keywords Paclitaxel · Tubulin · Binding pocket · MD simulation · DFT

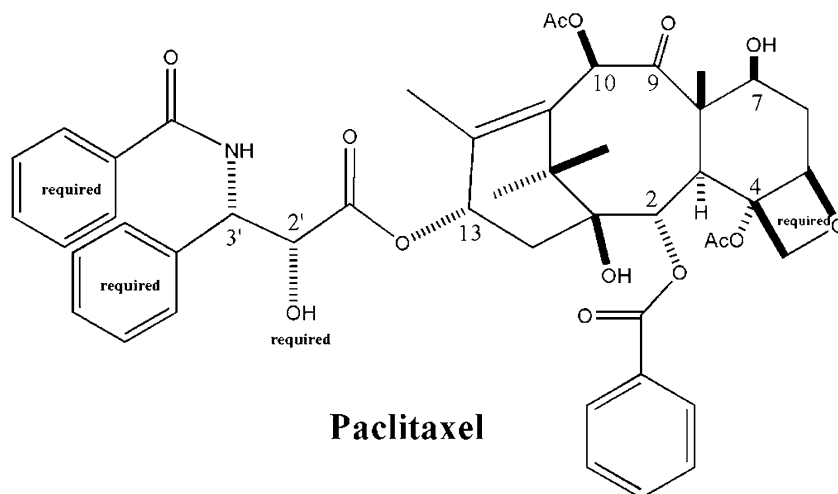
Introduction

Paclitaxel (PTX, Taxol), a diterpene compound (see Chart 1), is used to treat various cancers [1–3]. This drug has traveled a long path from its discovery to clinical use [4–9]. The clinical and commercial success of PTX has stimulated a worldwide search for compounds with a similar mode of action but superior properties, in particular reduced toxicity and enhanced activity against drug-resistant tumors [10]. As with many cancer chemotherapeutic agents, resistance and toxicity are still severe problems when treating malignancies with PTX [11, 12]. However, the exact mechanism by which human tumors become resistant to PTX is unknown. It is likely that several mechanisms occur in the active pocket during the binding of PTX to β -tubulin, which is associated with the clinical mechanism of PTX [13–35]. Currently, the principal clinical mechanism for PTX is considered to be the disruption of microtubule dynamics [36]. The normal functioning of tubulin assembly and disassembly is crucial to cell division. Any interference with this process will disrupt cell division and cause cell death by apoptosis [37]. When PTX binds to a microtubule at a site within the lumen, the dynamic properties of the microtubule are altered, leading to the stabilization of tubulin against depolymerization, which is necessary for cell division [38]. The binding of PTX within specific sites on β -

S. Xu · S. Chi · Y. Jin
Key Laboratory of Education Ministry for Medicinal Chemistry of
Natural Resource, College of Chemical Science and Technology,
Yunnan University,
Kunming 650091, People's Republic of China

S. Xu (✉) · Q. Shi · M. Ge · S. Wang · X. Zhang
State Key Laboratory for Structural Chemistry of Unstable and
Stable Species, Beijing National Laboratory for Molecular
Science, Institute of Chemistry, Chinese Academy of Sciences,
Beijing 100190, People's Republic of China
e-mail: sichuan@ynu.edu.cn

Chart 1 Paclitaxel (PTX) structure and atom numbering for the groups required for the groups required for activity against cancer cells



tubulin was observed in several photoaffinity labeling studies [39–42] and electron crystallographic (EC) structures [43–45]. The T-shaped conformation of PTX in its bound state was found to be the bioactive one, and the “southern hemisphere” of this conformation dominates in interactions with the active sites on β -tubulin [46, 47].

The key residues construct active compositions that form the PTX-binding pocket, although other residues may also mediate binding to PTX. When the key potential residues are identified on β -tubulin, we will gain a much clearer picture of the active pocket for PTX, which will then advance the study of the mechanism of resistance to PTX and promote the design of simpler analogs that are less toxic and are capable of avoiding resistance [41, 48, 49]. Accordingly, it is important to comprehend the PTX-binding pocket by identifying key residues that provide active sites within the surface of β -tubulin. In present work, we investigate the active pocket for the binding of PTX to tubulin by a docking method, molecular dynamics (MD) simulation, and density functional theory (DFT).

Materials and methods

β -Tubulin was obtained via separation from α,β -tubulin in the protein code 1JFF [43], and it contained 426 residues, one GDP and one PTX. Hydrogen atoms were added to it using the *Biopolymer* module [50] within the Sybyl software package (version 7.2) [51]. The positions of newly added hydrogen atoms were optimized using the *Minimize* module within Sybyl. Minimization was terminated at 0.01 kcal mol⁻¹ under the Tripos force field [51]. Residues within 6 Å of PTX were identified in order to extract single-residue–PTX pairs from the optimized hydrogen-added β -tubulin. Two hydrogen atoms were added to the carbonyl and amino terminals of each extracted residue for closure, as usual. These single-residue–PTX interacting pairs were not optimized further in order to maintain their original geometries. The binding energy (E_b) [52] for each extracted PTX–single-residue pair was calculated using B3LYP/6-31G(d,p) [53, 54] in the Gaussian 03 program [55] as follows:

$$E_b = E(\text{complex}) - E(\text{PTX}) - E(\text{residue}) \quad (1)$$

Table 1 E_b (kJ mol⁻¹), calculated with B3LYP/6-31G(d,p) for the EC-derived PTX, for each of the residues within 6 Å of the PTX bound to β -tubulin

Residue	E_b (BSSE)	Residue	E_b (BSSE)	Residue	E_b	Residue	E_b
Ala233	-7.0	Glu22	-14.5*(2.3)	Leu219	-1.3	Ser25	-0.9
Arg278	-30.9(6.4)	Glu27	-37.1(12.5)	Leu227	-0.1	Ser232	-1.0
Arg284	-12.7	Gly225	-0.5	Leu230	-0.7	Ser236	-4.0
Arg320	-0.2	Gly237	-4.2	Leu275	-9.0	Ser277	-3.0
Arg369	-16.9(10.9)	Gly279	-0.8	Leu371	-5.3	Ser374	0.6
Asp26	-19.4*(6.5)	Gly370	-12.3	Phe272	5.1	Thr216	0.6
Asp226	-20.5(4.9)	His229	-29.4*(14.0)	Pro274	-4.5	Thr234	-1.8
Cys213	-0.3	Ile24	-0.0	Pro359	-0.4	Thr276	11.2
Gln282	-2.5	Leu217	-0.7	Pro360	0.3	Val23	7.3
Sum							-216.8

* The only three key residues of active sites that were also identified by Jiang et al. in [21]

The Gromacs 3.3.3 software package was used to perform MD simulations [56–58]. Visualization of the simulation trajectories was achieved using the VMD software package [59]. The original extracted β -tubulin for the MD simulation contained 426 residues, one GDP and one PTX. The force-field parameters for GDP and PTX were defined individually, while those for β -tubulin were taken from the Gromacs package. For the GDP and PTX molecules, the gmx-united atom types were used, where the CH, CH₂ and CH₃ groups represented single atoms. Mulliken atomic charges (with hydrogens summed into heavy atoms) obtained using B3LYP/6-31G(d,p) were applied to the GDP and PTX molecules. Bond and angle parameters were defined according to the molecular structures of GDP and PTX optimized with B3LYP/6-31G(d,p). Improper dihedral parameters were defined in order to better strain the molecules into the appropriate configurations during simulation. Gromos96 force fields [60] were used in the MD simulations of β -tubulin.

The complex was immersed in an octahedral box of SPC waters [61, 62] during simulation, where there was a buffer of 2.0 nm between the protein atoms and the edge of the box. Na⁺ was introduced to neutralize the system. The

potential energy of the system was defined as being the sum of six contributions: chemical bonds, bond angles, bond twists, van der Waals, electrostatic, and hydrogen bonds. The protein geometry was first minimized until it satisfied the termination standard of 100 kJ mol⁻¹. The steepest descent energy minimization was used to remove bad van der Waals contacts between atoms. Next, the MD solute—whose equilibrium positions were restrained by the LINCS algorithm [63], while the solvent molecules were allowed to move in the simulation—was allowed to soak for 200 ps. The temperature was increased from 0 to 310 K under standard atmospheric pressure. Finally, the production simulation was performed for 10 ns with a time step of 2 fs; coordinates were written to the output file every 2000 steps (4 ps). The neighbor list, for which nonbonding interactions were calculated, was updated every 10 steps. Periodic boundary conditions were employed in the simulation box, along with a cutoff distance of 1.0 nm for electrostatic interactions and 1.0 nm for Lennard–Jones

Table 2 Various MD simulation times with energy minimization vs. $E(\text{LJ-SR})$ (kJ mol⁻¹) between PTX and β -tubulin and between PTX and Sol in the EC PTX–1JFF mode

MD time (ps)	PTX–tubulin	PTX–Sol
Crystal*	-347.2	-10.5
0	-358.3	-88.4
8	-364.1	-105.1
24	-368.8	-90.9
40	-377.7	-105.9
80	-399.7	-12.0
160	-424.9	-86.4
320	-395.1	-96.5
480	-397.6	-92.9
640	-379.5	-110.1
800	-378.1	-88.7
1000	-389.7	-85.1
2000	-376.4	-84.8
3000	-362.4	-101.9
4000	-340.6	-97.4
5000	-340.9	-114.3
6000	-346.2	-91.1
7000	-341.0	-112.6
8000	-330.8	-80.4
9000	-394.1	-89.6
10000	-359.7	-87.6

* $E(\text{LJ-SR}) = -304.7$ kJ mol⁻¹ for the EC-derived PTX–1JFF structure without energy minimization

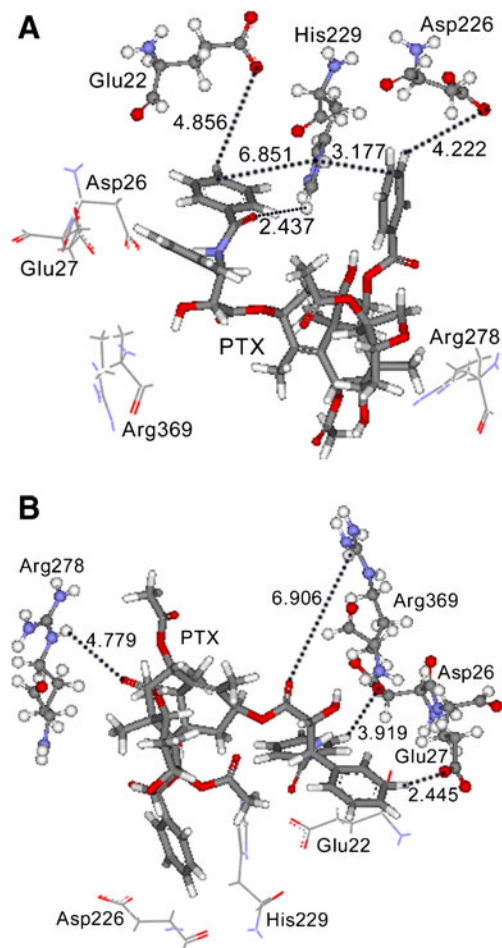


Fig. 1 The geometry of the EC-derived PTX when interacting with the key residues **a** Glu22, His229 and Asp226 and **b** Arg278, Arg369, Asp26 and Glu27 for the pure 1JFF EC complex. The atom distances (shown in Å) were measured using the Chem3D program

nonbonding interactions. The long-range electrostatic effect was treated with a PME procedure using default parameters. The Berendsen thermostat was used for a temperature of 310 K, along with a time constant of 0.1 ps. Berendsen exponential relaxation pressures coupled with a time constant of 0.5 ps were applied [64]. The trajectory data and frames written by the MD simulation were combined and saved into new files using the VMD program. The simulated protein coordinates were further minimized until they satisfied a strict termination standard of 10 kJ mol^{-1} with the same topology as used in the MD simulation. The energy groups were defined as protein, Sol, PTX and GDP in the minimization. The *g_energy* and *xmgrace* programs [56–58] were used to calculate and visualize the short-range Lennard–Jones interaction energies $E(\text{LJ-SR})$ between PTX and the protein and between PTX and Sol.

Two samples of β -tubulin, separated from the α,β -tubulins of the protein codes 1JFF and 1TVK [65], were used for docking assignments. In the 1TVK β -tubulin, the original ligand is an epothilone A molecule. In the 1JFF β -tubulin, the original ligand is an PTX molecule. Based on the conserved characteristics, the binding sites for docking PTX may be located at the same positions as those for the ligands. After a free-ligand β -tubulin had been prepared by removing the original ligand, PTX (which was separated from the 1JFF protein, as shown in Chart 1), was docked to the same binding sites as those for the ligands on the respective β -tubulin using the standard protocol of the docking program [66, 67]. The molecular docking process involves the prediction of the ligand conformation and its orientation within the protein interaction site. This is currently an increasingly important area of research in predictive structural chemistry, which has already yielded many publications [68–75]. In this work, the molecular surface of the binding site was calculated by means of software [67] to generate the Connolly surface with a probe of radius 1.4 Å. The points on the surface and the vectors normal to it were utilized by the Sphgen program [67] to build a set of spheres of radii 1.4–4 Å that can be used to visualize the site from a stereoelectronic point of view. The spheres within a radius of 10 Å from the reference ligand were used to represent the site. For each ligand, DOCK 6.0 calculated 500 orientations; the one with the best grid score was then considered. This grid-based score is based on the nonbonded terms of the molecular mechanical force field. The different PTX orientations relative to β -tubulin were ranked according to the minimized energies of the complexed structures. The ranked_PT_X orientation, which was normally specified as that corresponding to the best grid score as it was the minimum energy conformation, was combined with the respective β -tubulin using the Chimera program [76] and used for MD simulations, as described above.

Results and discussion

Currently, one of the most challenging problems in computational chemistry is to accurately calculate the binding of small molecule ligands to active sites on proteins [77–89]. Over the years, many different methods have been developed and used to model protein–ligand interactions, ranging from empirical modes to physics-based functions, with some even incorporating quantum mechanical calculations [90–99]. In the present work, Gromacs programs were used to perform calculations on the whole researched system, but the single-residue–ligand interactions were computed by pure quantum mechanics after the modeling. For each PTX–residue complex with more than one hundred atoms, E_b (which derives from internal energies, including weak intermolecular, electrostatic and hydrogen-bonding interactions) was calculated with the B3LYP method. The B3LYP functional takes thermochemical data [100, 101] that has been shown to be accurate for describing intermolecular and hydrogen-bonding interac-

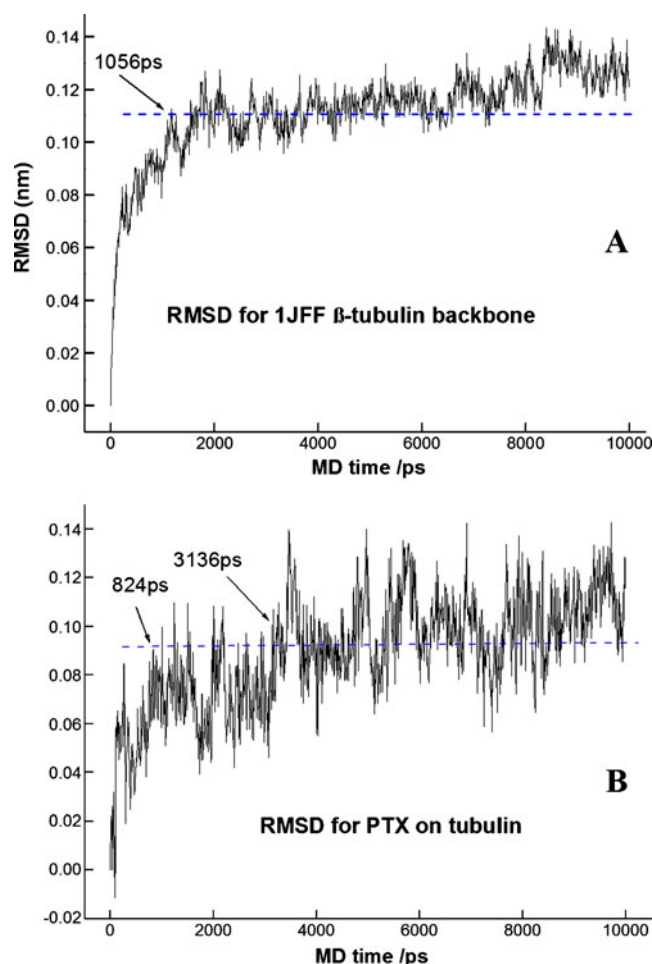


Fig. 2 Plots of RMSD (nm) vs. MD simulation time for the 1JFF β -tubulin backbone (a) and PTX (b) in the MD refined 1JFF EC mode

tions [102, 103], and uses it to evaluate intermolecular interactions of proteins, DNA and many other systems [104–110]. Table 1 lists the E_b values of each single-residue–PTX complex. The E_b values for PTX with Glu27, Arg278, His229, Asp226, Asp26, Arg369 and Glu22 are -37.1 , -30.9 , -29.4 , -20.5 , -19.4 , -16.9 and -14.5 kJ mol $^{-1}$, respectively. The basis set superposition error (BSSE), evaluated using the counterpoise procedure of Boys and Bernardi [111] with B3LYP/6-31G(d,p), is indicated in parentheses. Although it is fully corrected and sometimes super-estimated [112–114], the calculated BSSE is irregular and trivial. The values in Table 1 suggest that the BSSE is smaller if the interaction between PTX and the residue is weaker. In addition, because B3LYP calculates that less dispersion energy is integrated into E_b , it is better that the BSSE is not taken into account for E_b .

The E_b values of the single-residue–PTX complexes can be considered to lead to stable interactions of PTX with certain residues. Similarly, it is well known that chemical

reactions are characterized by an energy barrier arising from internal energy. The variation seen in $E(LJ-SR)$ for interactions of PTX with β -tubulin and Sol during simulation (Table 2) suggests that the solvation effect is much smaller for binding. Because the geometries of PTX–residue complexes were not further optimized with B3LYP/6-31G(d,p), they displayed fairly extended conformations, suggesting only very low structural entropy during binding. Even in the optimized systems, it was shown that the effects of entropy and temperature are small [115, 116]. However, to account for the possible effects of temperature, solvation and entropy, an E_b of -14.5 kJ mol $^{-1}$ (taken from the interaction of Glu22 with PTX) was defined as a boundary. The residue Glu22 is one of only three key residues (His229, Asp26, and Glu22) that were experimentally found by Jiang et al. to be potential sites for the binding of PTX to β -tubulin [21]. So far, only three experimentally found key residues have been reported for the interaction of β -tubulin with PTX, and no computationally found key

Table 3 E_b [kJ mol $^{-1}$, calculated with B3LYP/6-31G(d,p) and energy minimization] between PTX and each of the residues within 6 Å of PTX when bound to the EC 1JFF β -tubulin, where the complex

structure was sampled at various MD simulation times (as indicated by the values in parentheses)

Residue	$E_b(0.16$ ns)	$E_b(1$ ns)	$E_b(5$ ns)	Average	Residue	$E_b(0.16$ ns)	$E_b(1$ ns)	$E_b(5$ ns)	Average
Ala233	-1.4	-3.9	11.6	2.1	Leu227	0.2	0.8	-0.1	0.3
Ala367	-0.5				Leu230	-0.4	-3.2	-7.4	-3.7
Ala273		0.6			Leu275	-10.1	-0.2	-4.4	-4.9
Arg278	-87.9	-46.8	-68.5	-67.7	Leu371	-14.6	-16.9	-7.0	-12.8
Arg284	-19.1	-4.2	-15.4	-12.9	Lys372			-20.0	
Arg320	11.2	12.4	3.7	9.1	Lys19	8.2			
Arg369	-2.5	13.5	-15.8	-1.6	Lys218		-15.2		
Asn228	-0.7				Met365	1.3			
Asp26	-0.6	-25.3*	-3.6	-9.8	Phe20	-0.2			
Asp226	-24.3	-59.8	-63.6	-49.2	Phe272	0.2	-9.2	-0.2	-3.1
Cys213	1.3	-0.1	0.4	0.5	Pro274	-2.7	-1.8	5.5	0.3
Gln281			-14.5		Pro360	-0.8	4.6	-6.7	-0.9
Gln282	-7.7	-11.9	6.9	-4.2	Ser232	-7.3	-0.7	1.0	-2.3
Glu22	-23.2	-38.5*	-22.3	-28.0	Ser236	-2.2	-3.1	-4.1	-3.1
Glu27	-25.4	-54.6	-5.4	-28.5	Ser277	0.7	-0.9	0.9	0.2
Gly225	-0.6	-0.5	-2.2	-1.1	Ser374	-8.6	0.2	-3.6	-4.0
Gly237	-1.0	-2.6	5.6	0.7	Thr223	-1.3			
Gly279		0.3	0.4		Thr234	-0.7	1.3	-1.4	-0.3
Gly370	-11.8	-7.5	-7.3	-8.9	Thr276	-19.8	-16.5	-18.8	-18.4
His229	-70.0	-26.4*	-8.7	-35.0	Thr376	-1.1	-0.2	0.7	-0.2
Leu217	1.6	0.1	5.6	2.4	Tyr283	-0.3	-13.2	-6.1	-6.5
Leu219	-0.3	0.4	-0.8	-0.2	Val23	0.4	-1.7	3.5	0.7
Sum						-228.2	-330.7	-262.1	-273.7

Residue numbers were assigned such that they agree with those used in the EC mode. * Key residues of the active sites that were also identified by Jiang et al. [21]

residues on the protein have been reported. The exact boundary value of E_b , which is used to define which residues are actively involved in interactions with PTX, requires further study. In another work [117], which used the same boundary value of E , the active sites (residues) involved in interactions with PTX when binding to the reserved catalytic fragment of poly(ADP-ribose) polymerase were identified, and those residues were very similar to those identified experimentally [118, 119]. In that work, an effective method of searching for the key residues of active sites was validated [117]. Glu27, Arg278, His229, Asp226, Asp26, Arg369 and Glu22 were identified as the active

sites for interaction with PTX in the pure EC PTX–tubulin complex according to the E_b boundary defined above.

The geometry of PTX when interacting with the key residues can be clarified by labeling the distances shown in Fig. 1. Most of them contact with the “southern hemisphere” of PTX [120], which consists of C-13 to C-5 of PTX. Glu27, His229, Asp26, Arg369 and Glu22 interact with the C-13 hydroxyl chain of PTX. Arg278 lies adjacent to the “northern hemisphere” of PTX, which consists of C-7 to C-12. For the Asp and Glu residues, the $-\text{CO}_2^-$ group induces intermolecular interactions, so that the intermolecular distances involving this group are labeled. Glu27 faces

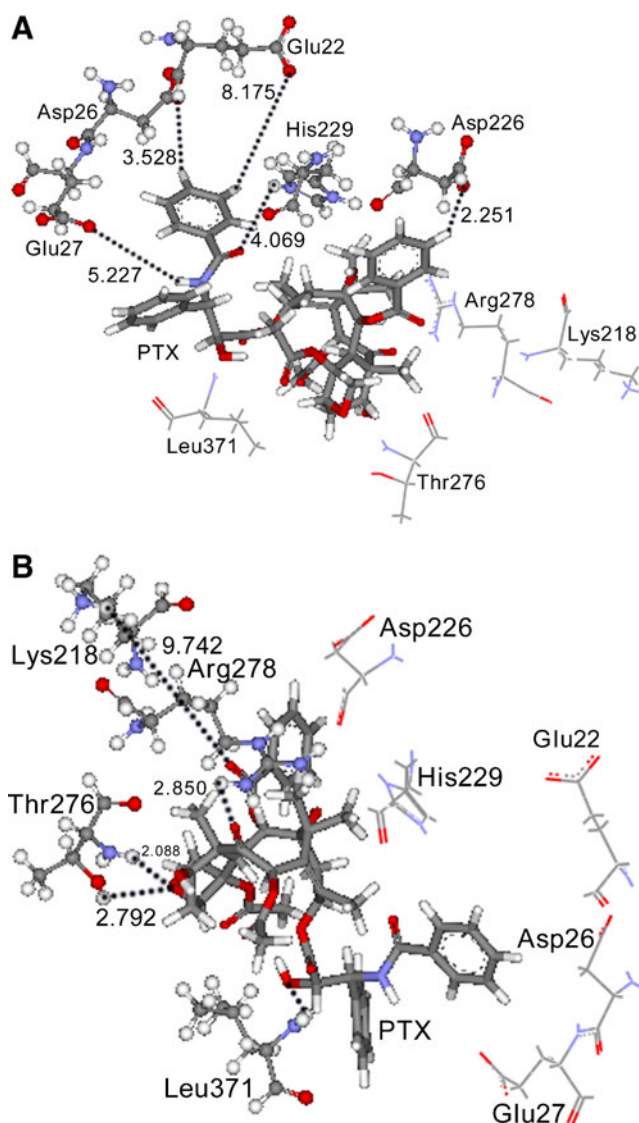
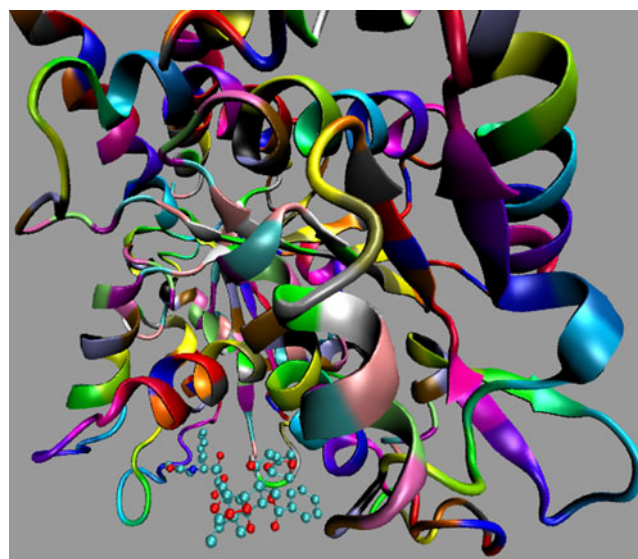
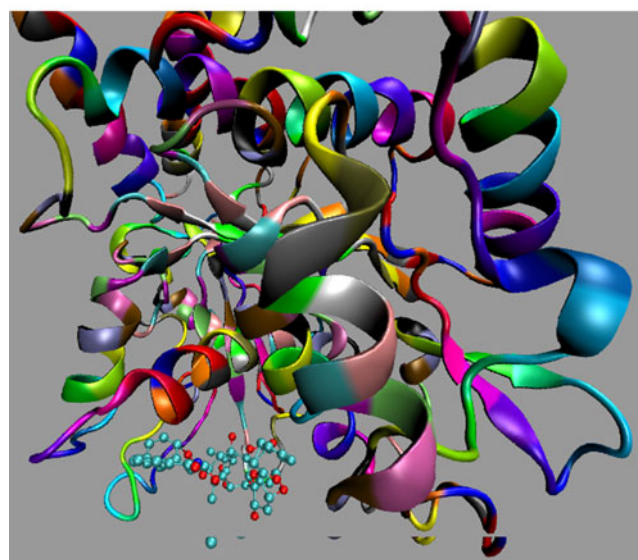


Fig. 3 The geometry of PTX when interacting with the key residues **a** Glu27, Asp26, Glu22, His229 and Asp226 and **b** Thr276, Lys218, Arg278 and Leu371 for the refined 1JFF EC complex from the 1 ns time point. Atom distances (shown in Å) were measured using the Chem3D program



PTX-1JFF tubulin (A)



PTX-1TVK tubulin (B)

Fig. 4 Two different orientations obtained when the PTX conformation is docked to the β -tubulins of 1JFF (a) and 1TVK (b)

the phenyl group of the C-3' chain, as shown by the closest distance of 2.445 Å between the O atom of -CO_2^- in Glu27 and the H atom of the phenyl of the C-3' chain, which is required for the activity of PTX against cancers [120]. Asp226 faces the acyloxy group of the C-2 chain, as suggested by the distance of 4.222 Å between the O atom of -CO_2^- in Asp226 and the H atom of the acyloxy group. Asp26 directs the N-acyl group, with a distance of 3.919 Å between the H atom of the N-acyl group and the O atom of -CO_2^- in Asp26. Glu22 also directs the N-acyl group, as suggested by the distance of 4.856 Å between the H atom of the N-acyl group and the O atom of -CO_2^- in Glu22. His229 is positioned between two phenyl groups at distances of 3.177 and 6.851 Å. The H atom of the $\equiv\text{NH}^+$ group in His229 in the natural state is responsible for electrostatic and hydrogen-bonding interactions with the O atom of the N-acyl group at a distance of 2.437 Å. Arg278 occupies a space near the C-9 and C-10 groups at a distance of 4.779 Å between the H atom in Arg278 and the O atom

of the C-10 (AcO-) group. Arg369, with one positive charge, faces the C-1' and C-2' groups of the C-13 chain, as shown by the distance of 6.906 Å between the H atom in Arg369 and the O atom of the C-1' group. Note that, if we consider the residues within 6 Å of PTX, some of the atomic distances between them are larger than 6 Å. From the geometry, it can be observed that the key residues use electrostatic, hydrogen-bonding and intermolecular interactions to interact with PTX.

Since the key residues are separated by peptide bonds, their interactions with PTX are characterized by orientation and saturation. The total E_b value between PTX and β -tubulin can be obtained by adding the E_b values between PTX and each residue within 6 Å of PTX. Short-range contacts between PTX and the key residues dominate the interactions. The total E_b value, $-216.8 \text{ kJ mol}^{-1}$ (see Table 1), represents the interaction between PTX and β -tubulin. In the MD calculation, it is $-304.7 \text{ kJ mol}^{-1}$ [from $E(\text{LJ-SR})$ between PTX and β -tubulin]. Because the

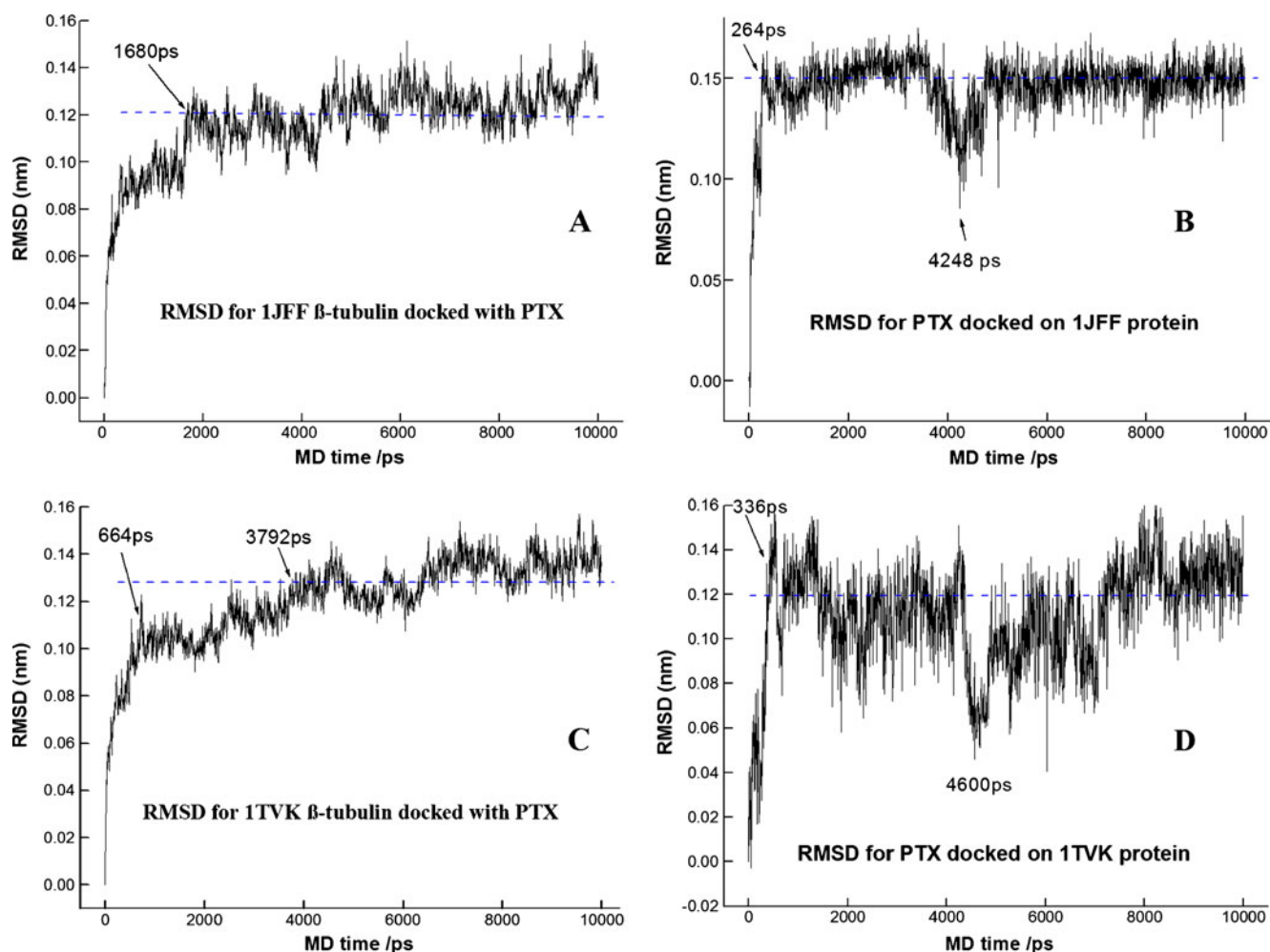


Fig. 5 Plot of RMSD (nm) vs. MD simulation time for: **a** 1JFF β -tubulin backbone docked with PTX, **b** PTX docked with 1JFF β -tubulin, **c** 1TVK β -tubulin backbone docked with PTX, **d** PTX docked with 1TVK β -tubulin

contents of the sphere within 10 Å of PTX are used to calculate $E(\text{LJ-SR})$, it is desirable to get a larger value in the MD calculation.

PTX is a highly flexible molecule due to its rotatable C-2, C-4, C-10, and C-13 groups, which are appended to an essentially rigid tetracyclic ring system, which raises the complicated question of how PTX interacts with protein residues [121]. However, an important achievement was the determination of the electron crystallographic structure of α,β -tubulin at a resolution of 3.5 Å [43]. Usually, the crystal protein structure is energy minimized, but an MD simulation is not performed. For the EC PTX–tubulin structure, an improved experimental density for the zinc-induced tubulin sheets was obtained using maximum

likelihood and Cartesian simulated annealing refinement. However, energy minimization alone was not able to produce a good fit to the 3D protein structure and adjust the interaction between PTX with tubulin appropriately. To better identify key residues, a 10 ns MD simulation of the EC-derived PTX–tubulin structure was required, during which the interaction energy $E(\text{LJ-SR})$ between PTX and β -tubulin was calculated in order to check the stability of their association, and the protein backbone RMSD was calculated to evaluate the stability of the protein. In Fig. 2a, the curve for the protein backbone RMSD shows that the MD simulation reaches a relatively stable equilibrium after 1056 ps. Meanwhile, as shown in Fig. 2b, PTX reaches equilibrium after 824 ps, as indicated by the stable PTX

Table 4 Various MD simulation times with energy minimization vs. $E(\text{LJ-SR})$ (kJ mol^{-1}) in the docked-PTX–1JFF mode

MD time (ps)	PTX–1JFF	PTX–Sol
8	–200.5	–172.2
16	–192.0	–161.1
24	–211.5	–153.0
32	–228.7	–150.0
40	–235.6	–140.5
48	–237.8	–141.6
56	–237.2	–144.9
64	–244.7	–138.3
72	–248.5	–125.3
80	–222.4	–129.0
88	–253.9	–123.2
168	–282.0	–116.3
248	–342.5	–99.6
328	–303.5	–106.6
408	–308.3	–96.7
488	–355.6	–83.4
568	–338.6	–89.7
648	–307.3	–84.6
728	–352.4	–85.8
808	–334.3	–73.5
888	–360.7	–81.2
968	–343.9	–99.6
1000	–335.5	–86.6
2000	–348.8	–77.8
3000	–351.5	–99.3
4000	–306.6	–102.0
5000	–315.4	–87.1
6000	–307.3	–105.2
7000	–313.8	–100.6
8000	–305.5	–94.5
9000	–315.2	–90.6
10000	–319.8	–108.9

Table 5 Various MD simulation times with energy minimization vs. $E(\text{LJ-SR})$ (kJ mol^{-1}) in the docked-PTX–1TVK mode

MD time (ps)	PTX–1TVK	PTX–Sol
8	–213.2	–142.7
16	–182.1	–121.8
24	–199.9	–129.8
32	–195.9	–110.5
40	–225.1	–138.3
48	–211.1	–129.0
56	–210.6	–148.3
64	–214.7	–150.0
72	205.7	–147.3
80	–220.8	–145.9
88	–228.5	–130.1
168	–306.5	–99.2
248	–324.4	–97.0
328	–292.0	–113.0
408	269.7	–111.3
488	–280.0	–117.2
568	–280.4	–108.4
648	–313.6	–111.8
728	–337.0	–123.9
808	–355.8	–99.1
888	–331.9	–103.7
968	–328.6	–122.4
1000	–316.4	–120.4
2000	–304.2	–102.5
3000	–316.0	–106.5
4000	–349.6	–92.5
5000	–305.0	–106.3
6000	–281.9	–113.8
7000	–322.4	–89.0
8000	–311.4	–106.2
9000	–299.3	–104.8
10000	–291.6	–100.6

RMSD fluctuations of ± 0.03 nm. In Table 2, fluctuations in the $E(\text{LJ-SR})$ data can be observed, as shown in Fig. 2, which may be caused by vibrations of the 3D structures of β -tubulin and PTX due to thermal motion.

To identify potential residues on β -tubulin that are responsible for PTX binding, three complex structures were extracted at 0.16, 1 and 5 ns during the MD trajectory and further refined by minimization. At the 0.16 ns time point, the complex produces the maximum interaction as shown by $E(\text{LJ-SR})$ (-424.9 kJ mol $^{-1}$). Table 3 lists the E_b values between individual residues and PTX within the three complexes, as calculated using B3LYP/6-31G(d,p). For certain residues, their E_b values with PTX were comparable

in the three structures, although the vibration of β -tubulin causes some discrepancies. It is known that a protein's function is determined by its three-dimensional structure and dynamics. Most proteins occur in unique conformations that are exquisitely well suited to their functions. In the functional conformation of β -tubulin, more residues likely serve as potential sites for PTX binding. Therefore, the refined complex from the 1 ns time point, which includes three experimentally identified key residues, was analyzed in detail. Based on the E_b boundary and values in Table 3, the residues Arg278, Asp26, Asp226, Glu22, Glu27, His229, Lys218, Leu371 and Thr276 were identified as potential sites for the binding of PTX refined

Table 6 E_b [kJ mol $^{-1}$, calculated with B3LYP/6-31G(d,p) and energy minimization] between PTX and each residue within 6 Å of PTX when bound to the 1JFF β -tubulin, where the complex structure was sampled at various MD simulation times (as indicated by the values in ns)

Residue	E_b						
	10 ns	5 ns	4 ns	3 ns	2 ns	0.888 ns	Average
Ala233	0.9	2.6	4.3	2.4	0.1	-4.3	2.06
Arg278	-127.9	-76.1	-76.1	-112.2	-141.5	-112.8	-106.8
Arg284	-5.9	-6.4	-16.7	-1.7	-9.9	-6.8	-8.1
Arg320	-12.7	-8.1	-5.6	-11.2	-14.6	-16.9	-10.4
Arg369	1.2	-18.4	-16.5	-20.3	-26.1	-26.8	-16.0
Asp26	-52.2	-37.5	-37.7	-64.1*	-69.4	-79.8	-52.2
Asp226	-20.1	-0.3	-26.4	-24.9	-18.6	-25.5	-18.1
Gln282	-0.5	-5.5	5.9	-1.0	-2.8	-4.0	-0.8
Glu22	-31.3	-34.2	-39.3	-38.1*	-46.5	-49.8	-37.9
Glu27	-17.2	-19.4	-25.1	-21.0	-29.1	-16.6	-22.4
Gly279			-2.2	-0.1	-2.9	-1.5	
Gly370	-0.0	0.6	1.6	0.9	-0.8	0.1	0.5
His28	-5.2	-10.1	-8.2				
His229	-69.5	-66.2	-9.2	-48.8*	-9.7	-12.7	-40.7
Ile24	0.7	1.1	1.2		0.6		
Leu217	1.5	0.5	4.3	-0.7	2.8	3.3	1.7
Leu230				-0.3	2.3	-0.3	
Leu275	-3.2	5.2	7.5	-1.3	-2.2	2.7	1.2
Leu286	0.1	-0.7	-0.4	-0.5	-0.4		-0.4
Leu371	-1.4	3.6	2.1	-3.3	2.7	0.4	0.7
Lys19					1.6		
Lys218	-9.2	-9.3	-6.5	-8.8	-11.0	-14.8	-9.0
Phe272	-4.0	-0.0	-2.1	-1.2	4.8	-1.6	-0.5
Pro274		-1.1	-7.2	5.6	0.9	-2.5	
Pro360	-2.0	-1.3	-0.9	-2.2	-3.1	-3.2	-1.9
Ser232	-1.4	0.4		0.1	-0.4	-1.5	
Ser236	-1.1	-0.5	-1.0		0.2	-1.8	
Ser277	-12.7	-9.3	-11.6	-16.8	-14.2	-13.9	-12.9
Ser374	7.7	-0.9	-0.5	-0.9	-0.3	-1.5	1.0
Thr216					2.2	3.4	
Thr234	0.1	0.6		0.7	-0.4	-0.5	
Thr276	7.3	1.3	-3.0	8.3	3.3	4.9	3.4
Val23	-0.5	-5.2	-1.3	-4.3	3.0	3.5	-1.7
Sum	-358.5	-294.6	-270.6	-365.7	-379.4	-355.4	-333.8

* Key residues of the active sites that were also identified by Jiang et al. in [21]

structure from the 1 ns time point. Among these, Leu371, Lys218 and Thr276 only serve as the potential sites in the refined complex sampled at 1 ns; they do not serve as active sites in the pure EC complex. The residue Arg369 serves as the active site in the pure EC complex, but is not a potential site in the refined complex, possibly due to thermal motions. The residues Arg278, Asp26, Asp226, Glu22, Glu27, and His229 are all potential active sites in the refined complex (from the 1 ns time point) and the pure EC mode.

In the refined structure from the 1 ns time point, Glu22, Asp26 and Glu27 lie adjacent to the phenyl group of the C-3' chain, as seen from the distances shown in Fig. 3. Asp226 and His229 contact with the acyloxy group of the C-2 chain. Thr276 interacts with the oxetane ring. Arg278 contacts with the O atom of the C-9 group through an electrostatic interaction. Leu371 faces the free 2'-hydroxyl group, with which it participates in a hydrogen-bonding interaction. Lys218 contacts with the O atom of $-C=O$ in the C-2 chain using an electrostatic interaction. Most of the key residues contact with the “southern hemisphere” of PTX, similar to what is seen in the pure EC mode. Only Arg278 interacts with the “northern hemisphere” of PTX in this refined mode.

It is possible that PTX actually uses different conformation orientations to bind to active sites within β -tubulin. This means that PTX would use more than one conformation orientation (unlike the EC mode) for binding, although this is not supported by the crystal protein structure. Figure 4 presents two distinct orientations of the PTX conformation obtained by docking PTX to the β -tubulins of 1JFF (4a) and 1TVK (4b). However, similar key residues can be identified in both conformation orientations as the active sites during interactions with PTX, described below.

The docked-PTX- β -tubulin structures were also investigated by MD simulation. In Fig. 5a, the curve for the β -tubulin backbone RMSD shows that the MD simulation reaches a relatively stable equilibrium for 1JFF β -tubulin docked with PTX after 1680 ps. Meanwhile, the docking of PTX to 1JFF β -tubulin reaches equilibrium after only 264 ps, as indicated by the stable PTX RMSD fluctuations of ± 0.03 nm illustrated in Fig. 5b. Similarly, the curve for the β -tubulin backbone RMSD shows that the MD simulation reaches a relatively stable equilibrium for 1TVK β -tubulin docked with PTX after 3792 ps, as shown in Fig. 5c. For PTX docked to 1TVK β -tubulin, it reaches equilibrium after 336 ps, as indicated by the stable PTX RMSD fluctuations of ± 0.03 nm in Fig. 5d. As listed in Tables 4 and 5, the interaction energy $E(LJ-SR)$ between β -tubulin and PTX decreases rapidly during the first nanosecond of MD simulation. After that, a relatively stable interaction between PTX and β -tubulin is achieved,

as suggested by $E(LJ-SR)$ values of -200.5 to -360.7 kJ mol $^{-1}$ for the docked-PTX-1JFF complex and -213.2 to -355.8 kJ mol $^{-1}$ for the docked-PTX-1TVK complex. Fluctuations in the $E(LJ-SR)$ data, caused by vibrations of the 3D structure of β -tubulin with PTX due to thermal motion, can also be observed in Tables 4 and 5.

Table 6 lists the E_b values (calculated using B3LYP/6-31G(d,p) and further refined by energy minimization) between individual residues and PTX within six complexes corresponding to the 0.888, 2, 3, 4, 5 and 10 ns time points of the MD trajectory. Certain residues show comparable E_b values with PTX across the six complexes, although the vibrations of β -tubulin cause some dissimilarity; the average results are listed in Table 6. In order to investigate the functional conformation of β -tubulin by evaluating which residues serve as potential sites for PTX binding

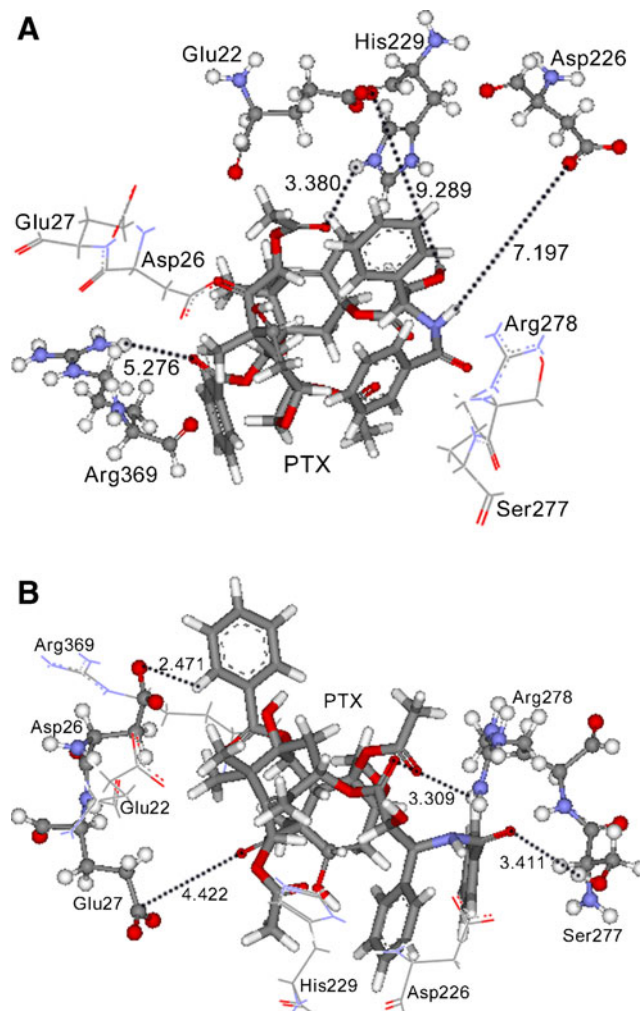


Fig. 6 The geometry of PTX when interacting with the key residues **a** Arg369, Glu22, His229 and Asp226 and **b** Glu27, Asp26, Arg278 and Ser277 for the refined docked 1JFF complex sampled at 3 ns. The atom distances (shown in Å) were measured using the Chem3D program

(including three experimentally identified key residues), the refined complex sampled at 3 ns was analyzed, and the results are shown in Table 6 and Fig. 6. Eight residues (Arg278, Arg369, Asp26, Asp226, Glu22, Glu27, His229 and Ser277) were identified as the active sites in the interaction with PTX based on their values in Table 6. The residue Ser277 is only a potential site in the refined complex corresponding to the 3 ns time point. The other seven residues (Arg278, Arg369, Asp26, Asp226, Glu22, Glu27, and His229) were all potential sites in both the

refined complex (for the 3 ns time point) or pure the EC mode.

Figure 6 shows the geometry of PTX when interacting with the key residues in the refined complex for the 3 ns time point. Arg278, Asp226, His229 and Ser277 contact with the C-13 chain. Arg278 and Ser277 face the the N-acyl group. Arg278 interacts with the O atom of the C-1' group. Ser277 faces the O atom of the C-3' chain. Asp226 and His229 direct the phenyl group, contacting with the H atom of –NH in the C-3' group and the O atom of the C-

Table 7 E_b [kJ mol⁻¹, calculated with B3LYP/6-31G(d,p) and energy minimization] between PTX and each residue within 6 Å of PTX when bound to 1TVK β -tubulin, where the complex structure was sampled at various MD simulation times (as indicated by the values in ns)

Residue	E_b					
	10 ns	5 ns	4 ns	3 ns	0.808 ns	Average
Ala233	0.1	3.5	6.2	4.6	-0.5	3.6
Arg278	-47.7	-30.6	-27.3	-7.1	-25.8	-28.2
Arg284	-6.7	-8.5	-9.0	-19.2	-14.0	-10.9
Arg320	-11.2	-10.0	-19.5	-18.0	-9.3	-14.7
Arg369	-3.4	0.4	-4.6	-7.9	-2.8	-3.9
Asp26	-30.1	-34.2*	-38.1	-37.1	-31.0	-34.9
Asp226	-33.9	-30.2	-45.7	-51.7	-35.3	-40.4
Cys213			1.1	0.9		
Gln282	-4.5		-6.0		-3.5	
Glu22	-50.5	-52.0*	-50.1	-41.5	-42.8	-48.5
Glu27	-14.4	-19.0	-16.9	-17.1	-18.9	-16.9
Gly279	-2.8	-22.2	-5.2	-2.5	-16.4	-8.2
Gly370	0.6	-3.7	-0.0	-5.3	-8.4	-2.1
His229	-30.1	-14.4*	-13.8	-5.2	-18.5	-15.9
Ile24			0.5	1.0		
Leu217	1.5	-1.1	-0.6	-5.9	1.8	-1.5
Leu219	-0.0	0.1	0.2	-2.0	0.2	-0.4
Leu230	1.1	-0.2	1.9	2.8	0.1	1.4
Leu275	-3.8	-8.4	-1.9	-8.2	-2.0	-5.6
Leu371	-3.3	-2.0	0.9	1.3	-1.1	-0.8
Lys19					13.1	
Lys218	-31.8	-31.0	-35.2	-37.2	-45.3	-33.8
Phe272	-1.9	2.0	-4.5	3.6	-6.6	-0.2
Pro274	-4.5	0.6	-3.9	3.5	8.2	-1.1
Pro360		0.3	-2.0	-7.2	-1.3	
Ser232	0.9				4.1	
Ser236					-1.2	
Ser277	-1.6	-21.7	-25.0	-26.5	-13.8	-18.7
Ser280		-6.9	-3.7	-4.9	-4.3	
Ser374	-1.1	-1.1	1.0	-6.0	-1.3	-1.8
Thr216				1.9	1.6	
Thr220	0.3	-1.2	-6.5			
Thr276	-29.3	-37.3	-32.0	-31.8	-1.5	-32.6
Val23	0.7	1.0	4.8	-1.1	0.4	1.35
Sum	-307.4	-327.8	-334.9	-323.8	-276.1	-323.5

* Key residues of active sites that were also identified by Jiang et al. [21]

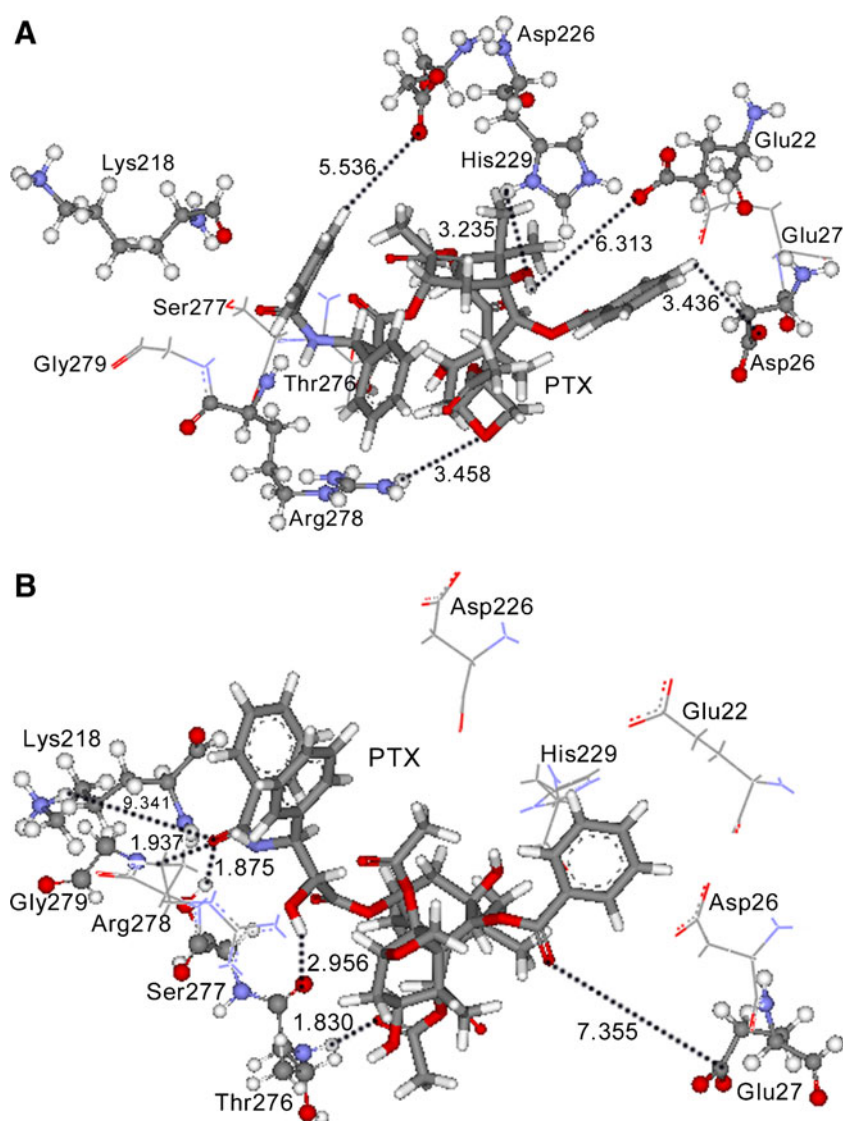
10 group. Arg369 and Asp27 both face the acyloxy group, contacting with the O atom of $-C=O$ in the C-2 group. Glu27 occupies a space near the C-9 and C-10 groups. Glu22 interacts with the free 2'-hydroxyl by long-distance electrostatic and hydrogen-bonding interactions. Except for Glu27, all of the key residues interact with the “southern hemisphere” of PTX within the refined complex.

Table 7 lists the E_b values between individual residues and PTX within five complexes sampled at 0.808, 3, 4, 5 and 10 ns during the MD trajectory and further refined by minimization. Certain residues show comparable E_b values with PTX across the five complexes, although the vibrations of β -tubulin also cause some differences; the average results are listed in Table 7. Again, in order to investigate the residues that serve as active sites that are responsible for interactions with PTX, including three experimentally

identified key residues, the refined complex sampled at 5 ns was analyzed in greater depth, and the results are shown in Table 7 and Fig. 7. Ten residues (Arg278, Asp26, Asp226, Glu22, Glu27, Gly279, His229, Lys218, Ser277 and Thr276) were found to be potential sites responsible for interactions with PTX, based on the values in Table 7. The residues Gly279 and Ser277 were potential sites in the refined complex for the 5 ns time point, but they were not active sites in the refined EC mode for the 1 ns time point. The residue Leu371 was a key residue in the refined EC mode from 1 ns, but was not a key residue in the refined dock mode from 5 ns. The other eight residues (Arg278, Asp26, Asp226, Glu22, Glu27, His229, Lys218 and Thr276) were all active sites in both modes.

Figure 7 shows the geometry of PTX when interacting with the key residues from the refined complex sampled at 5 ns. The residues Arg278, Asp226, Gly279, Lys218,

Fig. 7 The geometry of PTX when interacting with the key residues **a** Arg278, Asp226, His229, Glu22 and Asp26 and **b** Lys218, Gly279, Ser277, Thr276 and Glu27 for the refined docked 1TVK complex sampled at 5 ns. The atom distances (shown in Å) were measured using the Chem3D program



Ser277 and Thr276 all encompass the C-13 chain. Asp226 and Lys218 both face the N-acyl group of the C-3' chain. Gly279 and Ser277 direct the phenyl group, contacting with the O atom of the C-3' group and the H atom of –OH in the C-2' group. Arg278 faces the phenyl group, contacting with the O atom of the oxetane ring. Thr276 faces the phenyl group and contacts with the O atom of the C-10 group. Asp26, Glu22, Glu27 and His229 face the acyloxy group of the C-2 chain. Glu22 and His229 contact with the free OH of the C-1 group. Again, only one residue, Thr276, contacts with the “northern hemisphere” of PTX. The other nine key residues interact with the “southern hemisphere” of PTX within the refined complex sampled at 5 ns.

For the different PTX conformations, twelve residues were identified as the active sites in the binding of PTX to β -tubulin. The identified key residues, which can be divided into four groups, are shown in Fig. 8a. These residues construct a binding pocket within the interior surface of the β -tubulin, as shown in Fig. 8b. The first group consists of Glu22, Asp26 and Glu27. The second comprises Lys218, Asp226 and His229. The third consists of Thr276, Ser277, Arg278 and Gly279, while the fourth group is made up of Arg369 and Leu371. These key residues that form the active pocket in Fig. 8a work in a cooperative manner to enable PTX to bind to the interior surface of β -tubulin (Fig. 8b) in a strong interaction. The strength of this interaction is shown by the low total E_b value. For the refined EC complex sampled at 1 ns, the total E_b value was $-330.7 \text{ kJ mol}^{-1}$ (Table 3). For the

refined docked-PTX–1JFF complex sampled at 3 ns, the value was $-365.7 \text{ kJ mol}^{-1}$ (Table 6), and a value of $-327.8 \text{ kJ mol}^{-1}$ was obtained for the refined docked-PTX–1TVK complex sampled at 5 ns (Table 7). Even when PTX uses different conformations, similar interaction strengths between PTX and β -tubulin are noted. As demonstrated by these strength values, PTX can easily orientate and bind into the active pocket formed by the key active residues within the surface of β -tubulin.

Conclusions

We have identified that Arg278, Asp26, Asp226, Glu22, Glu27, His229, Arg369, Lys218, Ser277 and Thr276 are the active sites in the binding of PTX to β -tubulin. Another two residues, Leu371 and Gly279, also likely serve as active sites. Most of these residues interact with the “southern hemisphere” of PTX; only one contacts with its “northern hemisphere.” The residues can be divided into four groups, and they cooperatively form an active binding pocket within the interior surface of β -tubulin which is responsible for its interaction with PTX. The strength of this interaction is highlighted by the low total E_b values (ranging from -327.8 to $-365.7 \text{ kJ mol}^{-1}$) that result from different PTX conformations. Such a strong interaction allows PTX to easily bind into the active pocket, which in turn leads to the high activity of PTX against cancer cells.

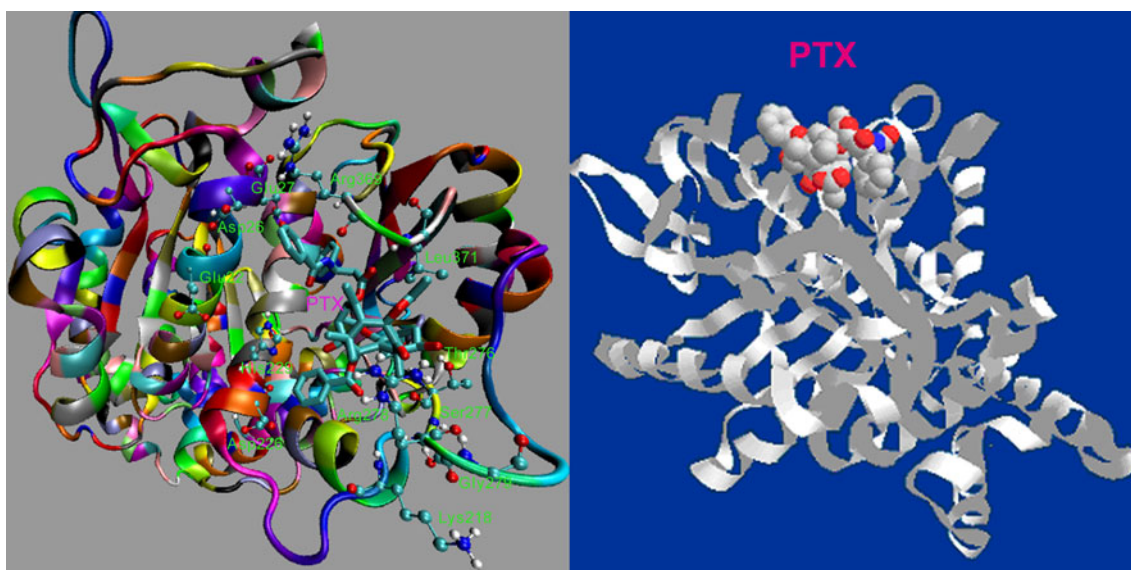


Fig. 8 **a** Stereo view of the active pocket formed by four groups of key sites in the binding of PTX to tubulin. The first group is Glu22, Asp26 and Glu27; the second group includes Lys218, Asp226 and His229; the third group consists of Thr274 Ser277, Arg278 and

Gly279; the fourth group is made up of Arg369 and Leu371. **b** PTX (shown as the *space-filling model*) binding to the active pocket located within the interior surface of β -tubulin (*ribbons*) in order to achieve a stable association with the tubulin

Acknowledgments The work was supported by One-Hundred-Talents Project of Chinese Academy of Sciences, and the Academic Talent Foundation of Yunnan Province, China (2006PY01-29).

References

- Brown DT (2003) *Taxus*: the genus *Taxus*. Taylor and Francis, London
- Eisenhauer EA, Vermorken JB (1998) *Drugs* 55:5–30
- Consumer Project on Technology (2010) Disputes involving Paclitaxel, a cancer drug sold under different brand names, including Taxol. <http://www.cptech.org/ip/health/taxol/>
- Giannakakou P, Snyder JP (2008) In: Fojo T (ed) Tubulin and microtubules. Humana, Totowa
- Goodman J, Walsh V (2001) The story of taxol: nature and politics in the pursuit of an anti-cancer drug. Cambridge University Press, Cambridge
- Kuppens IE (2006) *Curr Clin Pharmacol* 1:57–70
- Mooberry SL (2007) *Methods Mol Med* 137:289–302
- Schiff PB, Fant J, Horwitz SB (1979) *Nature* 277:665–667
- Scripture CD, Figg WD, Sparreboom A (2006) *Curr Neuropharmacol* 4:165–172
- Cowden CJ, Paterson I (1997) *Nature* 387:238–239
- Drukman S, Kavallaris M (2002) *Int J Oncol* 21:621–628
- Dumontet C, Sikic BI (1999) *J Clin Oncol* 17:1061–1070
- Alcaraz AA, Mehta AK, Johnson SA, Snyder JP (2006) *J Med Chem* 49:2478–2488
- Baker JK (1992) *Spectrosc Lett* 25:31–48
- Dubois J, Guénard D, Guéritte-Voegelein F, Guédira N, Potier P, Gillet B, Beloeil JC (1993) *Tetrahedron* 49:6533–6544
- Freedman H, Huzil JT, Luchko T, Luduen RF, Tuszynski JA (2009) *J Chem Inf Model* 49:424–436
- Geney R, Sun L, Pera P, Bernacki RJ, Xia S, Horwitz SB, Simmerling CL, Ojima I (2005) *Chem Biol* 12:339–348
- Hilton BD, Chmurny GN, Muschik GM (1992) *J Nat Prod* 55:1157–1161
- Hodge M, Chen QH, Bane S, Sharma S, Loew M, Banerjee A, Alcaraz AA, Snyder JP, Kingston DGI (2009) *Bioorg Med Chem Lett* 19:2884–2887
- Ivery TGM, Le T (2003) *Oncol Res* 14:1–19
- Jiang Y, Alcaraz AA, Chen JM, Kobayashi H, Lu YJ, Snyder JP (2006) *J Med Chem* 49:1891–1899
- Jiang Y, Lin HX, Chen JM, Chen MQ (2005) *Bioorg Med Chem Lett* 15:839–842
- Johnson SA, Alcaraz A, Snyder JP (2005) *Org Lett* 7:5549–5552
- Nogales E, Whittaker M, Milligan RA, Downing KH (1999) *Cell* 96:79–88
- Nogales E, Wolf SG, Khan IA, Luduena RF, Downing KH (1995) *Nature* 375:424–427
- Ojima I, Inoue T, Chakravarty S (1999) *J Fluorine Chem* 97:3–10
- Ojima I, Chakravarty S, Inoue T, Lin S, He L, Horwitz SW, Kuduk SD, Danishefsky SJ (1999) *Proc Natl Acad Sci USA* 96:4256–4261
- Ojima I, Kuduk SD, Chakravarty S, Ourevitch M, Bégue J-P (1997) *J Am Chem Soc* 119:5519–5527
- Paloma LG, Guy RK, Wrasidlo W, Nicolaou KC (1994) *Chem Biol* 1:107–112
- Querolle O, Dubois J, Thoret S, Roussi F, Guéritte F, Guénard D (2004) *J Med Chem* 47:5937–5944
- Shanker N, Kingston DGI, Ganesh T, Yang C, Alcaraz AA, Geballe MT, Banerjee A, McGee D, Snyder JP, Bane S (2007) *Biochemistry* 46:11514–11527
- Snyder JP, Nevins N, Cicero DO, Jasen J (2000) *J Am Chem Soc* 122:724–725
- Velde DGV, Georg GI, Grunewald GL, Gunn CW, Mitscher LA (1993) *J Am Chem Soc* 113:11650–11651
- Williams HJ, Scott AI, Dieden RA, Swindell CS, Chirlian LE, Francl MM, Heerding JM, Krauss NE (1993) *Tetrahedron* 49:6545–6560
- Wu Q, Bounaud PY, Kuduk SD, Yang CP, Ojima I, Horwitz SB, Orr GA (1998) *Biochemistry* 37:11272–11279
- Wang H, Nogales E (2005) *Nature* 435:911–915
- Kingston DGI (2009) *J Nat Prod* 72:507–515
- Jordan M, Wendell K, Gardiner S, Derry W, Copp H, Wilson L (1996) *Cancer Res* 56:816–825
- Combeau C, Commercon A, Mioskowski C, Rousseau B, Aubert F, Goeldner M (1994) *Biochemistry* 33:6676–6683
- Rao S, He L, Chakravarty S, Ojima I, Orr GA, Horwitz SB (1999) *J Biol Chem* 274:37990–37994
- Rao S, Krauss NE, Heerding JM, Orr GA, Horwitz SB (1994) *J Biol Chem* 269:3131–3134
- Rao S, Orr GA, Chaudhary AG, Kingston DG, Horwitz SB (1995) *J Biol Chem* 270:20235–20238
- Lowe J, Li H, Downing KH, Nogales E (2001) *J Mol Biol* 313:1045–1057
- Nogales E, Wolf SG, Downing KH (1998) *Nature* 391:199–203
- Snyder JP, Nettles JH, Cornett B, Downing KH, Nogales E (2001) *Proc Natl Acad Sci USA* 98:5312–5316
- Ganesh T, Guza RC, Bane S, Ravindra R, Shanker N, Lakdawala AS, Snyder JP, Kingston DGI (2004) *Proc Natl Acad Sci USA* 101:10006–10011
- Yang Y, Alcaraz AA, Snyder JP (2009) *J Nat Prod* 72:422–429
- Orr GA, Verdier-Pinard P, McDaid H, Horwitz SB (2003) *Oncogene* 22:7280–7295
- Geng X, Geney R, Pera P, Bernacki J, Ojima I (2004) *Bioorg Med Chem Lett* 14:3491–3494
- Clark M, Cramer RD, Opdenbosh NV (1989) *Comput Chem* 10:982–1012
- Tripos Inc. (2006) SYBYL Molecular Modeling System, 72nd edn. Tripos Inc., St Louis
- Xu SC, Deng S, Ma L, Shi Q, Ge M, Zhang X (2009) Active sites for retinal binding to bovine rhodopsin. *Acta Phys Chem Sin* 25:1290–1296
- Becke AD (1993) *J Chem Phys* 98:5648–5652
- Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37:785–789
- Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03, revision C.02. Gaussian Inc., Wallingford
- Berendsen HJC, Spoel DVD, Drunen RV (1995) *Comput Phys Commun* 91:43–56
- Spoel DVD, Lindahl E, Hess B, Buuren ARV, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers ALTM, Feenstra KA, Drunen RCv, Berendsen HJ (2005) Gromacs user manual, version 3.2. <http://www.gromacs.org>
- Spoel DVD, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) *J Comput Chem* 26:1701–1718
- Humphrey W, Dalke A, Schulten K (1996) *J Mol Graph* 14:33–38
- van Gunsteren WF, Billeter SR, Eising AA, Hunenberger PH, Krueger P, Mark AE, Scott WRP, Tironi IG (1996) *Biomolecular*

- simulation: the Gromos 96 manual and user guide, 1st edn. Hochschulverlag AG, ETHC Zurich, Zurich
61. Berendsen HJC, Postma JPM, Gunsteren WFv, Hermans J, Pullman B (2001) *J Am Chem Soc* 123:8638–8639
 62. Miyamoto S, Kollman PA (1992) *J Comput Chem* 13:952–962
 63. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) *J Comput Chem* 18:1463–1472
 64. Berendsen HJC, Postma JPM, Gunsteren WF, Dinola A, Haak JR (1984) *J Chem Phys* 81:3684–3690
 65. Nettles JH, Li H, Cornett B, Krahn JM, Snyder JP, Downing KH (2004) *Science* 305:866–869
 66. Ewing TJA, Kuntz ID (1997) *J Comput Chem* 18:1175–1189
 67. Lang PT, Moustakas D, Brozell S, Carrascal N, Mukherjee S, Pegg S, Raha K, Shivakumar D, Rizzo R, Case DA, Shoichet BK, Kuntz I (2006) DOCK 6.1. University of California, San Francisco
 68. Cambria MT, Di Marino D, Falconi M, Garavaglia S, Cambria A (2010) *J Biomol Struct Dyn* 27:501–510
 69. Da Cunha EFF, Barbosa EF, Oliveira AA, Ramalho TC (2010) *J Biomol Struct Dyn* 27:619–625
 70. Huang HJ, Lee KJ, Yu HW, Chen CY, Hsu CH, Chen HY, Tsai FJ, Chen CYC (2010) *J Biomol Struct Dyn* 28:23–37
 71. Nekrasov AN, Zinchenko AA (2010) *J Biomol Struct Dyn* 28:85–96
 72. Tao Y, Rao ZH, Liu SQ (2010) *J Biomol Struct Dyn* 28:143–157
 73. Kahlon AK, Roy S, Sharma A (2010) *J Biomol Struct Dyn* 28:201–210
 74. Tuccinardi T, Botta M, Giordano A, Martinelli A (2010) *J Chem Inf Model* 50:1432–1441
 75. Li Y, Shen J, Sun X, Li W, Liu G, Tang Y (2010) *J Chem Inf Model* 50:1134–1146
 76. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) *J Comput Chem* 25:1605–1612
 77. Faver JC, Benson ML, He X, Roberts BP, Wang B, Marshall MS, Kennedy MR, Sherrill CD, Merz KM Jr (2011) *J Chem Theor Comput* 7:790–797
 78. Gilson MK, Zhou HX (2007) *Annu Rev Biophys Biomol Struct* 36:21–42
 79. Leach AR, Shoichet BK, Peishoff CE (2006) *J Med Chem* 49:5851–5855
 80. Merz KM (2010) *J Chem Theor Comput* 6:1769–1776
 81. Dill KA (1997) *J Biol Chem* 272:701–704
 82. Hayik SA, Dunbrack R Jr, Merz KM Jr (2010) *J Chem Theor Comput* 6:3079–3091
 83. Lyne PD (2002) *Drug Discov Today* 7:1047–1055
 84. Jorgensen WL (2004) *Science* 303:1813–1818
 85. Abel R, Young T, Farid R, Berne BJ, Friesner RA (2008) *J Am Chem Soc* 130:2817–2831
 86. Moitessier N, Henry C, Maigret B, Chapleur Y (2004) *J Med Chem* 47:4178–4187
 87. Sutherland JJ, Nandigam RK, Erickson JA, Vieth M (2007) *J Chem Inf Model* 47:2293–2302
 88. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) *Nat Rev Drug Discovery* 3:935–949
 89. Lengauer T, Lemmen C, Rarey M, Zimmermann M (2004) *Drug Discov Today* 9:27–34
 90. Zhou T, Caflich A (2010) *Chem Med Chem* 5:1007–1014
 91. Raha K, Peters MB, Wang B, Yu N, Wollacott AM, Westerhoff LM, Merz KM (2007) *Drug Discov Today* 12:725–731
 92. Peters MB, Raha K, Merz KM (2006) *Curr Opin Drug Discov Dev* 9:370–379
 93. Raha K, Merz KM (2005) *J Med Chem* 48:4558–4575
 94. Fukuzawa K, Kitaura K, Uebayasi M, Nakata K, Kaminuma T, Nakano T (2005) *J Comput Chem* 26:1–10
 95. Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB (2002) *J Mol Graph Model* 20:281–295
 96. Bohm HJ (1998) *J Comput-Aided Mol Des* 12:309–323
 97. Wang R, Lu Y, Wang S (2003) *J Med Chem* 46:2287–2303
 98. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) *Proteins Struct Funct Genet* 52:609–623
 99. Trott O, Olson AJ (2009) *J Comput Chem* 31:455–461
 100. Capel-Sanchez MC, de la Peña-O’Shea VA, Barrio L, Campos-Martin JM, Fierro JLG (2006) *Top Catal* 41:1–4
 101. Cramer CJ, Tolman WB, Theopolid KH, Rheingold AL (2003) *Proc Natl Acad Sci USA* 100:3635–3640
 102. Mitra S, Singh TS, Mandal A, Mukherjee S (2007) *Chem Phys* 342:309–317
 103. Rao L, Ke HW, Fu G (2009) *J Chem Theor Comput* 5:86–96
 104. Główska ML, Martynowski D, Kozłowska K (1999) *J Mol Struct* 474:81–89
 105. Huang JS, Kertesz M (2007) *J Phys Chem A* 111:6304–6315
 106. Kim KS, Suh SB, Kim JC, Hong BH, Lee EC, Yun S, Arakeshwar P, Lee JY, Kim Y, Ihm H, Kim HG, Lee JW, Kim JK, Lee HM, Kim D, Cui C, Youn SJ, Chung HY, Choi HS, Lee CW, Cho SJ, Jeong S, Cho JH (2002) *J Am Chem Soc* 124:14268–14279
 107. Pablo PJD, Moreno-Herrero F, Colchero J, Herrero JG, Baro AM, Ordejon P, Soler JM, Artacho E (2000) *Phys Rev Lett* 85:4992–4995
 108. Sun CL, Wang CS (2009) *Chin Sci B* 39:481–487
 109. Szekeres ZS, Bogár F, Ladik J (2005) *Int J Quant Chem* 102:422–426
 110. Zhang SG, Zhang LC, Yang P (2008) *Acta Phys Chim Sin* 24:1637–1642
 111. Boys SF, Bernardi F (1970) *Mol Phys* 19:553–556
 112. Bene JED (1993) *J Phys Chem* 97:107–110
 113. van Duijneveldt FB, van de Rijdt JGCM, van Lenthe JH (1994) *Chem Rev* 94:1873–1885
 114. Kryachko ES, Zeegers-Huyskens T (2002) *J Phys Chem A* 106:6832–6838
 115. Xu SC, Ma LY, Bian FY, Shi Q, Ge MF, Zhang XK (2009) *Acta Phys Chim Sin* 25:2312–2318
 116. Xu SC, Deng SR, Ma LY, Shi Q, Ge MF, Zhang XK (2010) *Int J Quant Chem* 110:2671–2682
 117. Wang Y, Bian FY, Deng SR, Shi Q, Ge MF, Wang S, Zhang XK, Xu SC (2011) *J Biomol Struct Dyn* 28:881–893
 118. Ruf A, Mennissier de Murcia J, de Murcia G, Schulz GE (1996) *Proc Natl Acad Sci USA* 93:7481–7485
 119. Ruf A, de Murcia G, Schulz GE (1998) *Biochemistry* 37:3893–3900
 120. Kingston DGI (2000) *J Nat Prod* 63:726–734
 121. Ganesh T, Yang C, Norris A, Glass T, Bane S, Ravindra R, Banerjee A, Metaferia B, Thomas SL, Giannakakou P, Alcaraz AA, Lakdawala AS, Snyder JP, Kingston DGI (2007) *J Med Chem* 50:713–725

Computational study of the electronic structures, UV-Vis spectra and static second-order nonlinear optical susceptibilities of macrocyclic thiophene derivatives

Shuang Huang · Ai-Min Ren · Lu-Yi Zou · Yang Zhao ·
Jing-Fu Guo · Ji-Kang Feng

Received: 21 February 2011 / Accepted: 4 April 2011 / Published online: 3 May 2011
© Springer-Verlag 2011

Abstract Using thiophene (which has a moderate resonance energy) as a spacer rather than benzene permits better π -electron delocalization and leads to a large nonlinear optical response. Thus, the nonlinear optical coefficients of a series of macrocyclic thiophene derivatives (C[3T_DA]_n with C_n symmetry) were studied, and their electronic structures, UV-Vis spectra and static second-order nonlinear optical susceptibilities (β_0) were computed. The calculated results showed that ΔE_{H-L} increased and the UV-Vis spectrum redshifted as the number of C[3T_DA] units increased (one C[3T_DA] unit consists of trithiophene and diacetylene). The value of β_0 calculated by either the ZINDO-SOS or the FF method showed the same trend: the absolute value of β_0 increased as the number of units increased. The value of β_0 predicted by ZINDO-SOS was an order of magnitude larger than that predicted by the FF method. However, the results suggest that macrocyclic thiophene compounds potentially exhibit large static second-order nonlinear optical susceptibilities.

Keywords Macrocyclic thiophene derivatives · Static second-order nonlinear optical susceptibilities · ZINDO-SOS method · FF method

Introduction

In recent years, it has been shown that aromatic-based systems such as oligophenylene exhibit limited chemical and photothermal stabilities due to the excessive confinement of the π -electrons associated with the high aromatic stabilization energy of the benzene ring. In this regard, π -conjugated spacers [1–5] based on thiophene have progressively emerged as an interesting tradeoff that allow the elaboration of nonlinear optical (NLO) compounds with $\beta\mu$ values (the scalar product of the vector component of the first hyperpolarizability tensor and the dipole moment) which sometimes exceed 10^{-44} esu. These properties are related to the moderate resonance energy of thiophene, which permits better π -electron delocalization than obtained with benzene [6].

There is contemporary research interest in the development of second-order nonlinear optical materials exhibiting large β values for possible applications in electro-optic devices. Recently, many novel molecules have been synthesized, including aromatic and heterocyclic compounds. The nonlinear optical (NLO) properties of these novel compounds have been studied both theoretically [7, 8] and experimentally [9, 10]. These investigations suggest that the molecular geometries of most nonlinear optical materials with large second-order nonlinear responses usually involve a chain, a branched geometry, or a

S. Huang · A.-M. Ren (✉) · L.-Y. Zou · Y. Zhao · J.-K. Feng
State Key Laboratory of Theoretical and Computational
Chemistry, Institute of Theoretical Chemistry, Jilin University,
Changchun 130023, People's Republic of China
e-mail: aimin_ren@yahoo.com

J.-F. Guo (✉)
School of Physics, Northeast Normal University,
Changchun 130021, People's Republic of China
e-mail: guojf217@nenu.edu.cn

spherical shape [11–13]. However, large π -electron systems of closed rings have not been investigated. Previous research [14] indicated that compounds with large π -conjugated structures possess many attractive NLO characteristics. Therefore, in the reported here, we focused on the relationships between electronic structure and NLO properties and between molecular size and NLO properties in a series of macrocyclic thiophene derivatives.

We focused on a series of $C[3T_DA]_n$ ($n=2-5$) compounds (Fig. 1) that consisted of two, three, four and five repeated $C[3T_DA]$ units, respectively. The $C[3T_DA]$ unit consists of trithiophene and diacetylene. These four macrocyclic molecules were previously synthesized, and showed high fluorescence quantum yields and improved fluorescence lifetimes [15–17], but their second-order nonlinear optical properties were not reported. Therefore, our research mainly concentrated on the static second-order nonlinear optical properties of this series of $C[3T_DA]_n$ ($n=2-5$) macrocycles with C_n symmetry.

In order to obtain the static second-order nonlinear optical susceptibilities (β_0) and the structure– β_0 relationship for these macrocycles, as well as to explore the origin of their second harmonic generation, their electronic structures and UV-Vis spectra were determined using Truhlar's MPWB95 functional [18, 19], which is based on the modified Perdew and Wang 1991 exchange functional [20] (mPW or MPW) and Becke's 1995 meta correlation functional [21] (B95). Then their static second-order nonlinear optical susceptibilities were calculated by the ZINDO method (ZINDO = Zerner's intermediate neglect of differential overlap) [22] combined with the sum-over-states (SOS) expression, and by the finite field (FF) method on the basis of the optimized geometries. The results showed that the studied molecules possess rela-

tively large β_0 values, which may originate from the effect of the π -conjugation on the annular skeleton, through the analysis of the electron cloud distribution and bond length alternation (BLA). In order to investigate the different characteristics of the annular compounds and their corresponding linear structures, the β_0 values of four linear molecules, $C[3T_DA]_n$ -L ($n=2-5$; for their optimized geometries, see Fig. 2), were calculated by the ZINDO-SOS and FF methods, respectively.

Theoretical methods

A quantitative description of the molecular NLO response can be derived from a power-series expansion of the polarization of the molecule upon its interaction with an external electromagnetic field, as shown in the following equation:

$$P_i = \sum_j \alpha_{ij} E_j + \sum_{j \geq k} \beta_{ijk} E_j E_k + \sum_{j \geq k \geq l} \gamma_{ijkl} E_j E_k E_l + \dots \quad (1)$$

Here, P_i is the molecular polarization induced along the i^{th} axis; E_j is the j^{th} component of the applied electric field; α , β and γ are the linear polarizability, the first hyperpolarizability or the second-order polarizability, and the second hyperpolarizability or the third-order polarizability, respectively.

α , β and γ describe the susceptibility of the molecule to an electromagnetic perturbation, and are constants for a given molecular geometry and external electromagnetic field. β and γ are responsible for second-harmonic and third-harmonic generation. The SOS expression for the

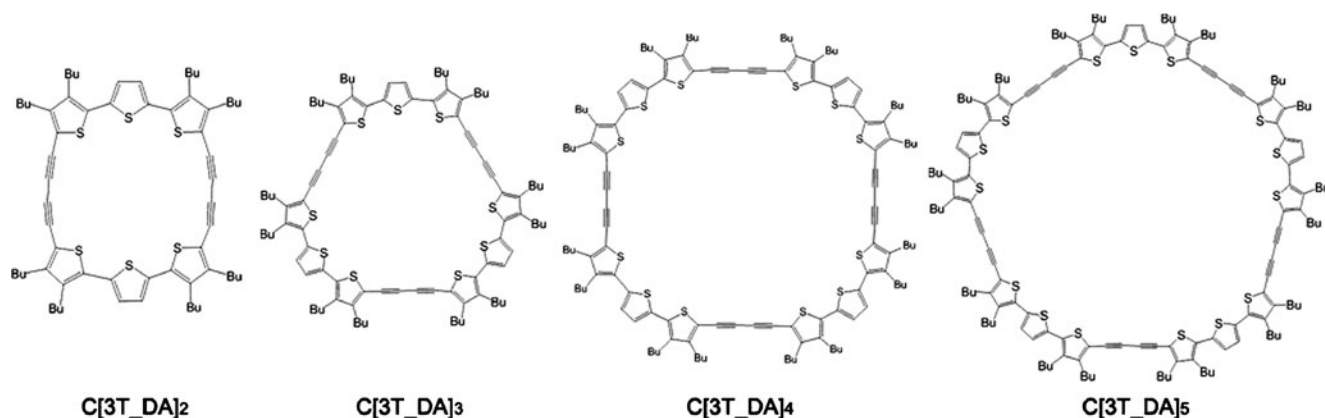


Fig. 1 Structures of the molecules with C_n symmetry studied in this work

individual components of the second-order polarizability tensor is obtained from perturbation theory [23–27]:

$$\beta_{ijk} + \beta_{ikj} = -\frac{e^3}{4\hbar^2} \left\{ \sum_{\substack{n \neq n' \\ n \neq g \\ n' \neq g}} \left[\begin{aligned} & \left(r_{gn'}^j r_{n'n}^i r_{gn}^k + r_{gn'}^k r_{n'n}^j r_{gn}^i \right) \times \left(\frac{1}{(\omega_{n'g} - \omega)(\omega_{ng} + \omega)} + \frac{1}{(\omega_{n'g} + \omega)(\omega_{ng} - \omega)} \right) + \\ & \left(r_{gn'}^i r_{n'n}^j r_{gn}^k + r_{gn'}^k r_{n'n}^i r_{gn}^j \right) \times \left(\frac{1}{(\omega_{n'g} - 2\omega)(\omega_{ng} - \omega)} + \frac{1}{(\omega_{n'g} + 2\omega)(\omega_{ng} + \omega)} \right) + \\ & \left(r_{gn'}^j r_{n'n}^k r_{gn}^i + r_{gn'}^k r_{n'n}^j r_{gn}^i \right) \times \left(\frac{1}{(\omega_{n'g} - \omega)(\omega_{ng} - 2\omega)} + \frac{1}{(\omega_{n'g} + \omega)(\omega_{ng} + 2\omega)} \right) \end{aligned} \right] + \right. \\ \left. 4 \sum_{n \neq g} \left[r_{gn}^j r_{gn}^k \Delta r_n^i (\omega_{ng}^2 - 4\omega^2) + r_{gn}^i \left(r_{gn}^k \Delta r_n^j + r_{gn}^j \Delta r_n^k \right) (\omega_{ng}^2 + 2\omega^2) \right] \frac{1}{(\omega_{ng}^2 - \omega^2)(\omega_{ng}^2 - 4\omega^2)} \right\} \quad (2)$$

where the summations are performed over the complete sets of eigenstates $|n\rangle$ and $|n'\rangle$ of the unperturbed molecular system. The quantities $r_{gn'}^i (\langle g|r^i|n\rangle)$ and $r_{nn'}^i (\langle n|r^i|n'\rangle)$ are matrix elements of the position operator r^i (along the i^{th} coordinate axis) between the unperturbed ground state (g) and excited states (n), and between the two excited states (n, n'), respectively; $\Delta r_n^i = r_{nn}^i - r_{gg}^i$ is the dipole moment difference between the excited state (n) and ground state (g). Although all 27 components of the β tensor can be computed, only the vector component in the dipolar direction (β_μ) is sampled by electric-field-induced second-harmonic generation experiments. Also, β_μ is given by

$$\beta_\mu = (\mu_x \beta_x + \mu_y \beta_y + \mu_z \beta_z) / (\mu_x^2 + \mu_y^2 + \mu_z^2)^{1/2}, \quad (3)$$

where

$$\beta_i = \beta_{iii} + 1/3 \sum_{i \neq j} (\beta_{jii} + \beta_{iji} + \beta_{ijj}) i, j \in (x, y, z). \quad (4)$$

The ZINDO program was employed to provide the transition dipole moment and the transition energy needed in Eq. 2 in the next step. The reliability of the ZINDO-SOS method for describing molecular NLO properties has been demonstrated previously [24, 28]. The second-order NLO polarizabilities [29–31] and the third-order NLO polarizabilities [32, 33] of dozens of molecules were reported in our previous work [12, 34]. In this study, the configuration interaction (CI) includes singlet excitations generated from the 20 highest occupied and 20 lowest unoccupied orbitals, and doublet excitations

from the five highest occupied and five lowest unoccupied orbitals, giving a total of 826 configuration states in the ZINDO computation.

There is a truncation problem during the computation of β with the SOS (Eq. 2). The calculated β value will reach a limiting value as the number of the excited states increases. It was shown that a summation over 25–400 excited states is usually adequate to achieve the convergence of the β value [35, 36]. In this study, convergence of the calculated β was realized by summing over 290 excited states.

In the SOS formula for β_{ijk} (Eq. 2), if we only consider an important excited state that possesses a low-energy excited state with high oscillator strength, then combining Eq. 3 and Eq. 4 yields the following approximate two-level formula [25, 26, 37], which is still a good model to use to obtain a chemical interpretation of the NLO response:

$$\beta_\mu = \frac{3e^2 \hbar^2}{2m} \frac{Wf \Delta\mu}{[W^2 - (2\hbar\omega)^2][W^2 - (\hbar\omega)^2]}. \quad (5)$$

Here, ω is the frequency of the laser field and $W = \hbar\omega_{1g}$ is the transition energy from the ground state to the excited state; f is the oscillator strength of the transition; $\Delta\mu$ is the difference in the dipole moments of the excited state and the ground state. β_μ can also be expressed as

$$\beta_\mu = \frac{\beta_0 W^4}{[W^2 - (2\hbar\omega)^2][W^2 - (\hbar\omega)^2]}, \quad (6)$$

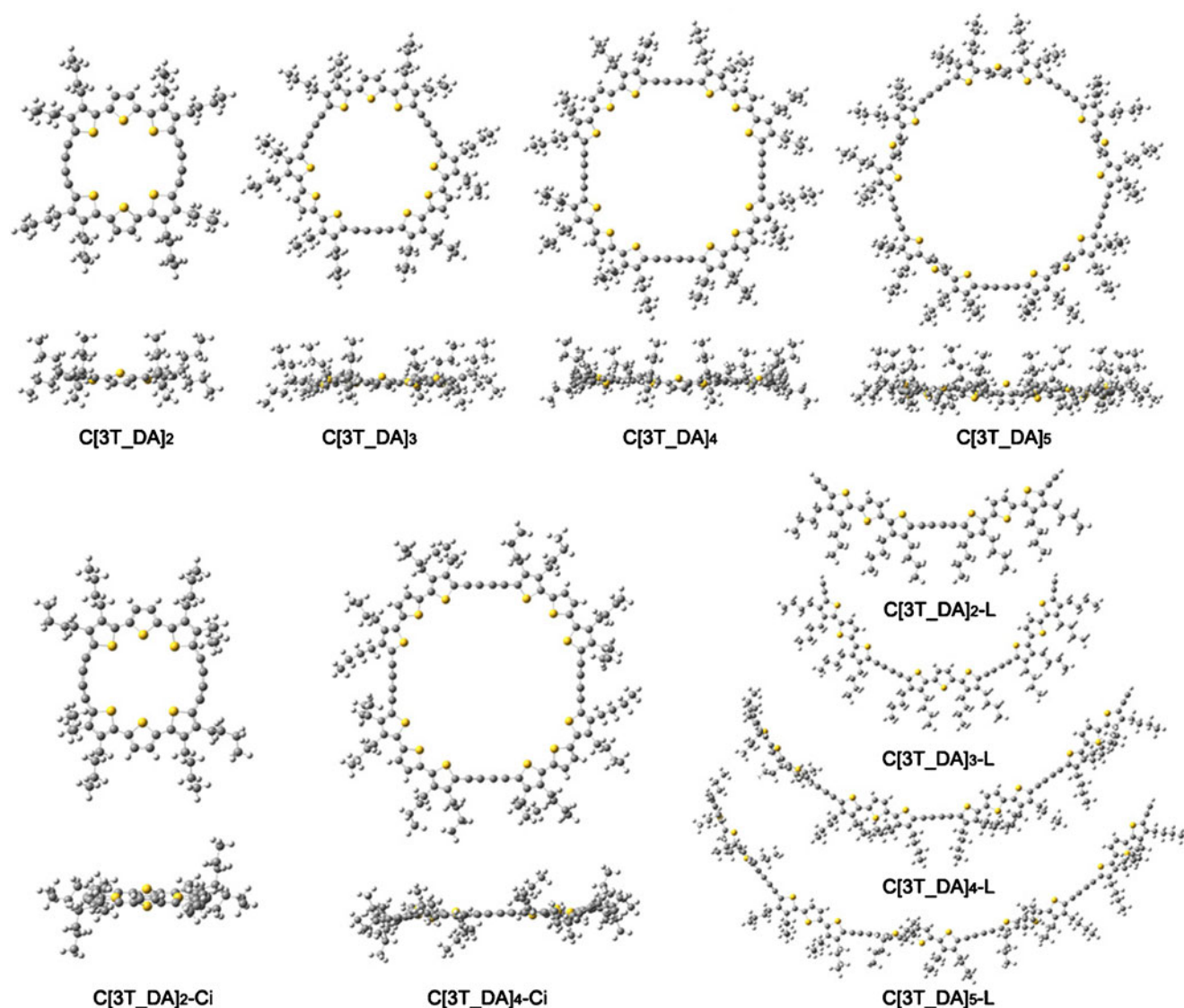


Fig. 2 Optimized geometries of $C[3T_DA]_n$ and $C[3T_DA]_2-L$

where

$$\beta_0 = \frac{3e^2 \hbar^2 f \Delta\mu}{2mW^3} = B \frac{f \Delta\mu}{W^3}. \quad (7)$$

β_0 is known as the static second-order nonlinear optical susceptibility. It is also termed the zero-frequency hyperpolarizability, and is an estimate of the intrinsic molecular hyperpolarizability in the absence of resonance effects.

Besides the ZINDO-SOS method, which has proven to be reliable for describing molecular NLO properties, the finite field (FF) method is also broadly used to compute nonlinear optical coefficients. Hence, we calculated the β_0 values of all of the studied molecules using DFT theory in Gaussian 03 [38] for the sake of comparison and to verify the trend in β_0 calculated using the ZINDO-SOS method. The output from Gaussian 03 provides ten components of

the $3 \times 3 \times 3$ matrix: β_{xxx} , β_{xxy} , β_{xyy} , β_{yyy} , β_{xxz} , β_{xyz} , β_{yyz} , β_{xzz} , β_{yzz} , and β_{zzz} , respectively [39, 40], from which the x , y and z components of β are calculated via Eq. 3. This is known as the finite field (FF) method. Since these β values of Gaussian 03 are reported in atomic units (a.u.), the calculated total values are converted into electrostatic units (esu) ($1 \text{ a.u.} = 8.6393 \times 10^{-33} \text{ esu}$). In the ZINDO-SOS method, molecular orbitals are obtained by INDO-SCF calculations, while ground and excited states are obtained through CI calculations. The FF method adds an interaction term to the Hamilton relating to the external electric field and charge (electron, atomic nucleus) [41]. NLO response computations based on ab initio model Hamiltonians or DFT Hamiltonians with the FF prescription for computing β have provided numerically accurate responses.

Table 1 UV-Vis maximum absorption peaks (λ_{\max}), oscillator strengths (f) and transition characteristics of C[3T_DA] $_n$, as calculated by the ZINDO and TD MPWB95/6-31 G* methods

Molecule	ZINDO		TD MPWB95/6-31 G*				
	λ_{\max} (nm)	f	λ_{\max} (nm)	f	Transition characteristics		
C[3T_DA] $_2$	409.4	0.963	401.0	1.208	$S_0 \rightarrow S_4$	HOMO \rightarrow LUMO+1	27.07%
			400.5 ^a			HOMO-1 \rightarrow LUMO	6.75%
	392.5	1.649	378.4	0.993	$S_0 \rightarrow S_6$	HOMO \rightarrow LUMO+2	41.59%
C[3T_DA] $_3$	421.4	2.392	374.4 ^a	2.659	$S_0 \rightarrow S_2$	HOMO \rightarrow LUMO+1	21.22%
			400 [15–17]			HOMO-2 \rightarrow LUMO	10.03%
						HOMO-1 \rightarrow LUMO	7.56%
	421.4	2.392	433.5	2.659	$S_0 \rightarrow S_3$	HOMO \rightarrow LUMO+2	21.22%
						HOMO-1 \rightarrow LUMO	10.03%
						HOMO-2 \rightarrow LUMO	7.56%
C[3T_DA] $_4$	452.1	3.011	422 [16, 17]	3.707	$S_0 \rightarrow S_2$	HOMO \rightarrow LUMO+1	20.94%
			447.8			HOMO-2 \rightarrow LUMO	17.62%
			432.6 ^a			HOMO-1 \rightarrow LUMO	17.62%
	452.1	3.011	447.8	3.707	$S_0 \rightarrow S_3$	HOMO \rightarrow LUMO+2	20.94%
			415.2 ^a			HOMO-1 \rightarrow LUMO	17.62%
C[3T_DA] $_5$	464.0	3.592	420 [16, 17]	4.725	$S_0 \rightarrow S_2$	HOMO \rightarrow LUMO+2	18.56%
			458.1			HOMO-1 \rightarrow LUMO	18.48%
	464.0	3.592	458.1	4.726	$S_0 \rightarrow S_3$	HOMO \rightarrow LUMO+1	18.56%
						HOMO-2 \rightarrow LUMO	18.48%
			433 [15]				
			431 [16, 17]				

^a Calculated for C_i symmetry

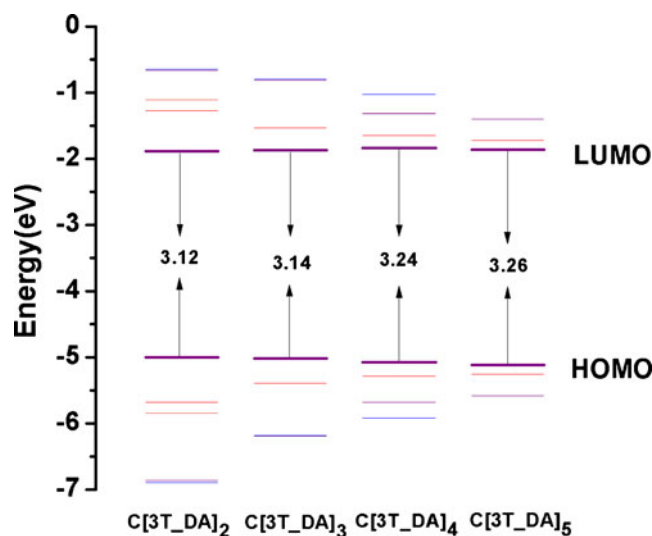
Results and discussion

Geometry optimization

The molecular structures of the C[3T_DA] $_n$ ($n=2-5$) compounds studied in this research (depicted in Fig. 1) were fully optimized at the MPWB95 level with the 6-31 G* basis set. All of the optimized geometries (Fig. 2) have C_n ($n=2-5$) symmetry, while C[3T_DA] $_5$ also has C_{5V} symmetry [42]. In addition, both C[3T_DA] $_2$ and C[3T_DA] $_4$ have other isomers with C_i symmetry [43, 44]. It is well known that the second-order nonlinear optical coefficient (β) of a molecule with C_i symmetry is zero [45], so isomers of C[3T_DA] $_2$ and C[3T_DA] $_4$ with C_i symmetry were not considered. In this work, we focused solely on macrocyclic thiophenes with axial symmetry (C_n) for our systematic comparison.

The optimized results show that the main body moiety of each molecule is not planar, and the arrangement of the middle thiophene in each unit differs from those of the other two, and it turns in the opposite direction to them. The two butyls on each thiophene in the molecules with C_n symmetry distort in two directions in order to

balance the overall structure. The diacetylene group that connects two trithiophene units shows enhanced molecular coplanarity.

**Fig. 3** Predicted (using MPWB95/6-31 G*) molecular orbital energy diagram

UV-Vis spectra

The UV-Vis spectra were calculated using the TD MPWB95/6-31 G* method on the basis of the geometries optimized via the MPWB95/6-31 G* method. The calculated UV-Vis absorption wavelength maxima (λ_{\max}), the corresponding oscillator strengths (f), and the transition characteristics are listed in Table 1. Among the results, the calculated λ_{\max} values of C[3T_DA] $_n$ ($n=2-5$) with C_n symmetry are in good agreement with the experimental values (400 nm, 422 nm, 420 nm, 433 nm or 431 nm) [15–17]. The UV-Vis maximum absorption peaks of C[3T_DA] $_2$ and C[3T_DA] $_4$ with C_i symmetry were also calculated, and are shown in Table 1. Both of these λ_{\max} values are smaller than that obtained for C_n symmetry, but the overall trend remains the same as above: the λ_{\max}

values of C[3T_DA] $_n$ ($n=2-5$) increase with increasing unit number. Moreover, it is known that molecules with C_i symmetry have no second-order nonlinear optical response since their dipole moments are zero, so such molecules will not be considered in the following discussion. Comparing the experimentally determined absorption band maxima with theoretical excitations, the agreement between them gives us the confidence to make further theoretical predictions for this macrocyclic system.

Using the optimized geometries obtained by the MPWB95 method, UV-Vis spectra were calculated by the ZINDO method. During the calculation, the $\pi-\pi$ overlap weighting factor was adjusted to be 0.51 based on TDDFT results and experimental values. Thus, the maximum absorption peaks were calculated to occur at 392.5 nm, 421.4 nm, 452.1 nm and 464.0 nm, respectively

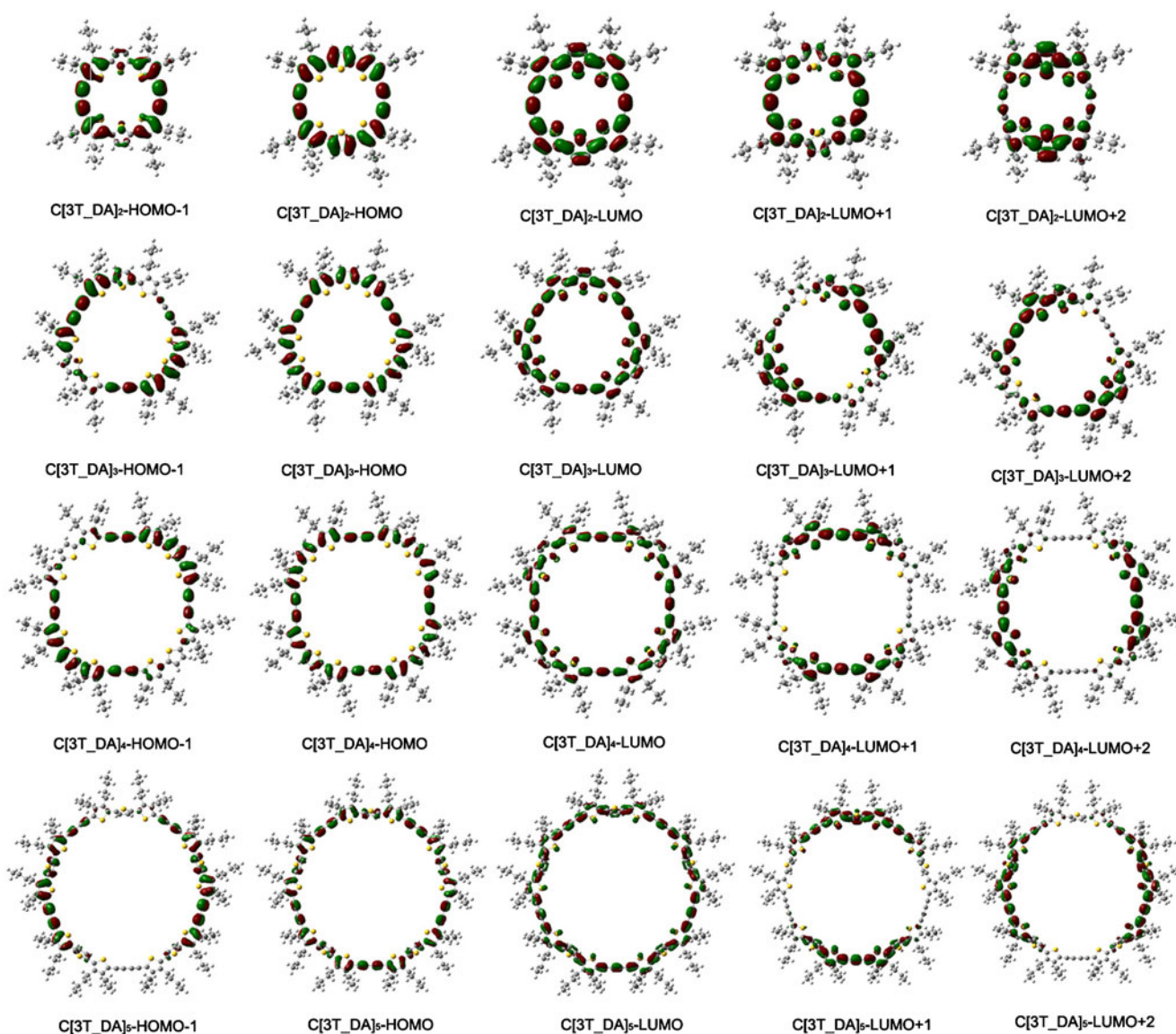


Fig. 4 Contour surfaces of the frontier orbital of C[3T_DA] $_n$

Table 2 The net charges of the ground and main excited states of the main moieties in C[3T_DA]_n

Molecule		Ground state (q_g)			Excited state (q_e)			Δq		
		-C≡C- C≡C-	Lateral thiophene	Middle thiophene	-C≡C- C≡C-	Lateral thiophene	Middle thiophene	-C≡C- C≡C-	Lateral thiophene	Middle thiophene
C[3T_DA] ₂	S ₀ →S ₄	-0.2586	-0.0193	-0.1540	-0.2451	-0.0244	-0.1662	0.0135	-0.0051	-0.0122
	S ₀ →S ₅	-0.2586	-0.0193	-0.1540	-0.2090	0.0171	-0.2464	0.0496	0.0364	-0.0924
C[3T_DA] ₃	S ₀ →S ₂	-0.4046	-0.0201	-0.2270	-0.3830	-0.0272	-0.2470	0.0216	-0.0071	-0.0200
	S ₀ →S ₃	-0.4046	-0.0201	-0.2270	-0.3830	-0.0272	-0.2470	0.0216	-0.0071	-0.0200
C[3T_DA] ₄	S ₀ →S ₂	-0.5473	-0.0237	-0.2971	-0.5211	-0.0445	-0.3086	0.0262	-0.0208	-0.0115
	S ₀ →S ₃	-0.5473	-0.0237	-0.2971	-0.5211	-0.0445	-0.3086	0.0262	-0.0208	-0.0115
C[3T_DA] ₅	S ₀ →S ₂	-0.7021	-0.0490	-0.3748	-0.6663	-0.0829	-0.3820	0.0358	-0.0339	-0.0072
	S ₀ →S ₃	-0.7021	-0.0490	-0.3748	-0.6663	-0.0829	-0.3820	0.0358	-0.0339	-0.0072

(Table 1). As shown in Table 1, C[3T_DA]₃, C[3T_DA]₄ and C[3T_DA]₅ all have two degenerate states, the second (S₂) and the third excited states (S₃); that is, the calculated transitions are from the ground state (S₀) to S₂ or S₃. Also, the corresponding oscillator strength obviously increases as the number of C[3T_DA] units increases. Therefore, for

this series of macrocycles C[3T_DA]_n ($n=2-5$), when the unit number increases and the annulus gets bigger, the absorption band is increasingly redshifted, and the oscillator strength is enhanced. All of these phenomena can be attributed to the enlargement of the π -conjugated region.

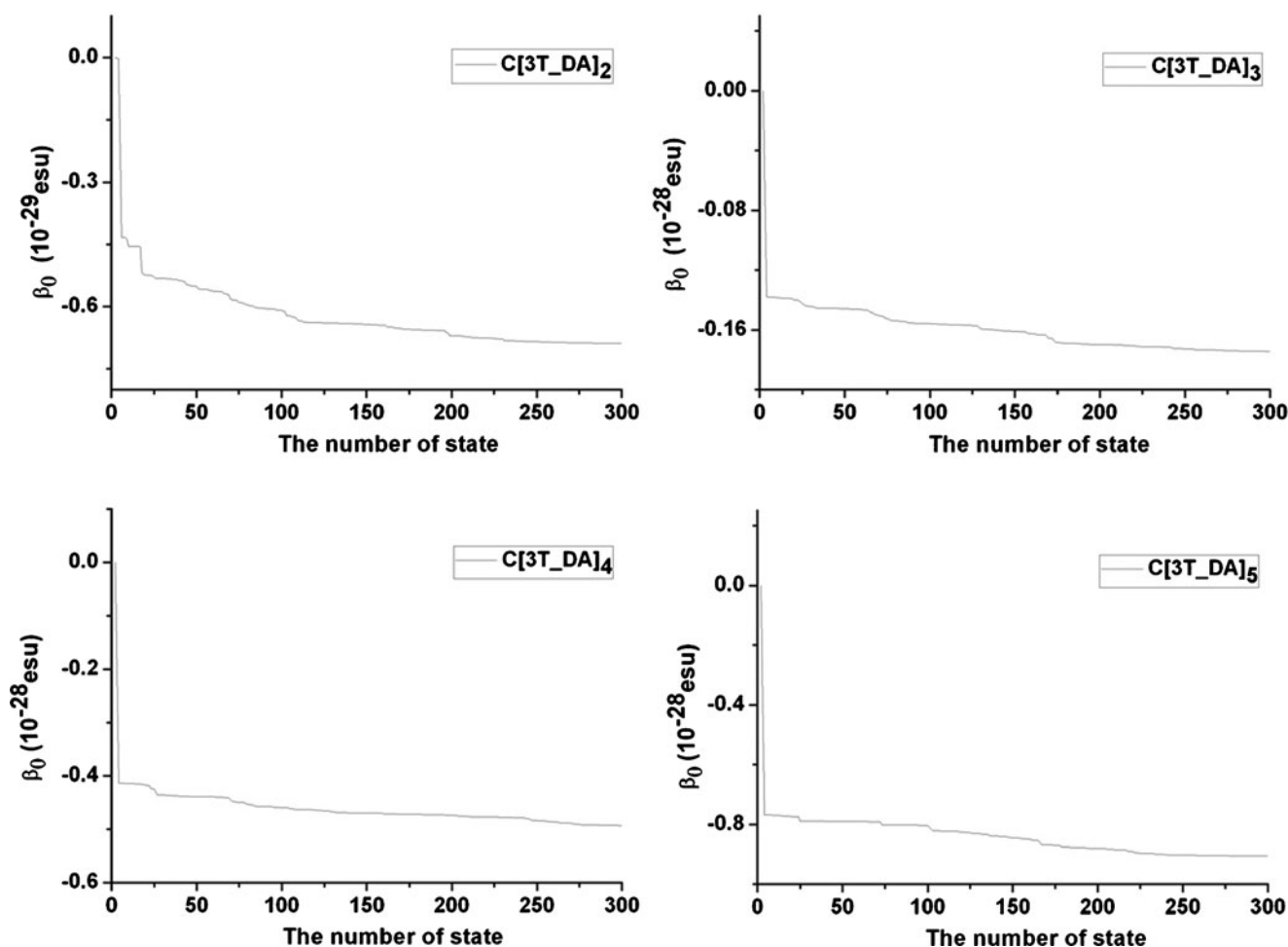
**Fig. 5** The relationship between β_0 and the number of excited states for C[3T_DA]_n, obtained with the ZINDO-SOS method

Table 3 The parameters of C[3T_DA]_n and C[3T_DA]_n-L calculated with the ZINDO-SOS method

ZINDO-SOS	N_π	$S_0 \rightarrow S_k$	f	W (cm ⁻¹)	μ_0 (debye)	μ_k (debye)	$\Delta\mu$ (debye)	$\beta_0 \times 10^{-30}$ (esu)
C[3T_DA] ₂	16	$S_0 \rightarrow S_5$	1.649	25476.4	0.950	0.939	-0.011	-6.88
C[3T_DA] ₃	24	$S_0 \rightarrow S_3$	2.392	23732.2	1.425	1.269	-0.156	-17.43
C[3T_DA] ₄	32	$S_0 \rightarrow S_3$	3.011	22119.6	2.752	2.774	0.023	-49.24
C[3T_DA] ₅	40	$S_0 \rightarrow S_3$	3.592	21551.5	3.925	3.963	0.038	-90.52
C[3T_DA] ₂ -L	16	$S_0 \rightarrow S_2$	2.303	22154.0	7.060	6.213	-0.847	-77.61
C[3T_DA] ₃ -L	24	$S_0 \rightarrow S_1$	3.194	22024.6	4.824	5.072	0.248	-73.60
C[3T_DA] ₄ -L	32	$S_0 \rightarrow S_1$	3.601	21415.1	10.167	10.281	0.114	-260.10
C[3T_DA] ₅ -L	40	$S_0 \rightarrow S_1$	4.393	21922.9	10.318	10.415	0.097	-345.55

Electronic structure and charge transfer

In order to compare the energy gaps of these complexes, the molecular HOMOs (the highest occupied orbitals) and LUMOs (the lowest unoccupied orbitals) were examined. The molecular orbital energies of C[3T_DA]_n (with C_n symmetry) calculated by the MPW1B95 method are depicted in Fig. 3, which shows that the energy of the HOMO decreases and the energy of the LUMO increases as the number of units increases, so that ΔE_{H-L} (the energy gap between the HOMO and the LUMO) also increases. However, λ_{\max} also increases, which may be attributed to the important contributions from the corresponding excited state transitions originating from some higher energy excitations, such as HOMO \rightarrow LUMO+1 and HOMO \rightarrow LUMO+2 (Table 1), but not HOMO \rightarrow LUMO, probably because of molecular symmetry. Furthermore, we focused on the ΔE_{H-L+1} (energy gap between HOMO and LUMO+1) and ΔE_{H-L+2} (energy gap between HOMO and LUMO+2) values. Figure 3 indicates that both ΔE_{H-L+1} and ΔE_{H-L+2} decrease as the unit number increases, which will lead to a redshifted of λ_{\max} . It is evident that the trend in the energy gap between the HOMO and the LUMO does not always correspond to that of the absorption wavelength maximum.

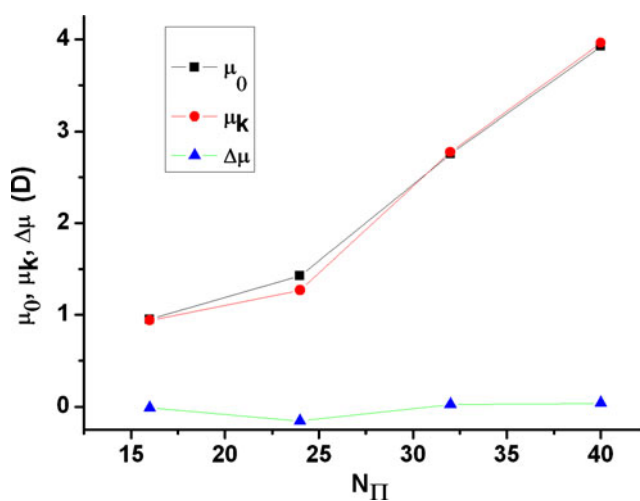
As is well known, the molecular orbitals are closely related to the electronic spectrum, so the molecular orbitals related to the main transitions are shown in Fig. 4. It is easy to see that all of the electron clouds in these orbitals are located on the big ring, while none of them are on the ambient butyl. In other words, the π -conjugated rings provide significant contributions to the HOMOs. The two excited singlet configurations HOMO \rightarrow LUMO+1 and HOMO \rightarrow LUMO+2 are strongly mixed, while the distributions of electron clouds on LUMO+1 and LUMO+2 are complementary. Transitions between the HOMOs and LUMOs accompany obvious intramolecular charge transfer in the macrocyclic skeleton for this series of molecules.

In order to analyze intramolecular charge transfer in C[3T_DA]_n ($n=2-5$), the net charges in the ground and main excited states on the main moieties, such as the acetylene

bridge and the lateral and middle thiophenes, were calculated and are listed in Table 2. The value of Δq is defined as the difference in charge between the ground state and the excited state ($\Delta q = q_e - q_g$), where q_g denotes the charge population in the ground state and q_e is the charge population in the excited state. Table 2 shows that the absolute values of Δq for the acetylene and lateral thiophene moieties increase as the number of units increases, except in the case of C[3T_DA]₂, and the value of the middle thiophene moiety decreases. In other words, charge transfer mainly occurs between the acetylene and the thiophene moieties. Overall, there is a great deal of charge transfer across the main body skeleton.

NLO properties

The NLO properties of the studied molecules were investigated based on the optimized geometries and the correct electronic spectra predicted by the TDDFT and ZINDO methods. It is well known that the ZINDO method combined with the SOS expression can accurately predict the second-order nonlinear optical coefficient (β), including

**Fig. 6** The dependences of μ_0 , μ_k and $\Delta\mu$ on N_π

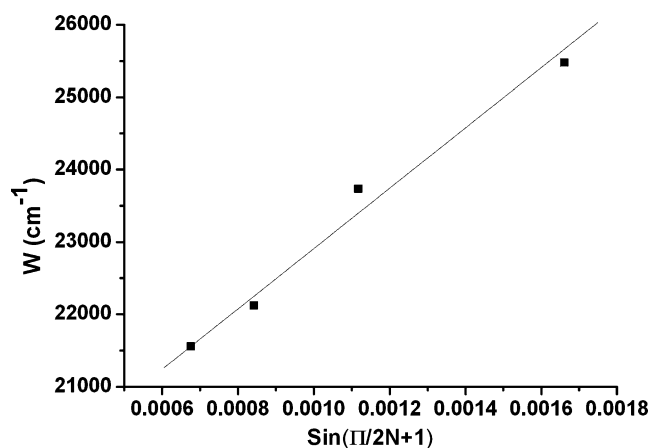


Fig. 7 The dependence of W on $\sin[\pi/(2N+1)]$

the static one (β_0) [13, 29]. Here, we first calculated β_0 using this method. In the SOS expression (Eq. 2) for β , the summation is performed over the complete set of eigenstates $|n\rangle$, $|m\rangle$ and $|g\rangle$ of the unperturbed system. In Eq. 2, the denominator of each term includes ω_{ng} and $\omega_{n'g}$. With increasing n and n' , the energy differences between the ground state and the excited states will increase, and the denominator increases. This means that as the energy of the excited state increases, its contribution to the β value decreases. When n and n' reach a certain value, the higher excited states are neglected. Li et al. [13] demonstrated that 197 excited states were found to be sufficient to permit effective convergence for C_{60} . However, for the macrocyclic thiophene derivatives in our study, 300 excited states were considered when calculating β_0 . From Fig. 5, it is clear that the values of β_0 gradually decrease as the number of excited states increases. The β_0 converge when the number of excited states exceeds 290. The values of β_0 and related parameters calculated in this way for $C[3T_DA]_n$ molecules are presented in Table 3. It shows that the

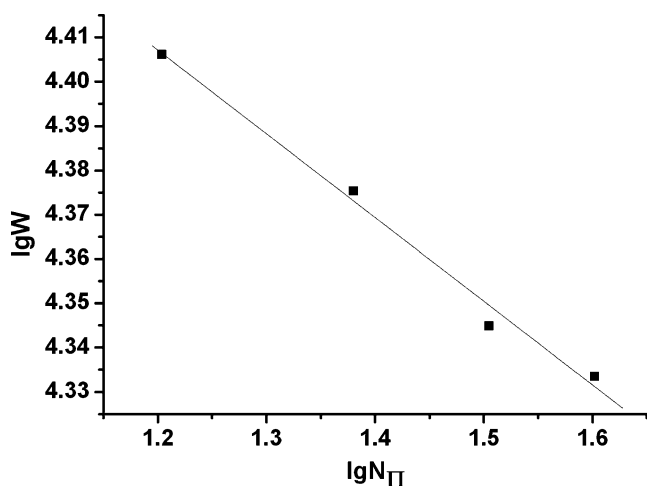


Fig. 8 The dependence of $\lg W$ on $\lg N_\pi$

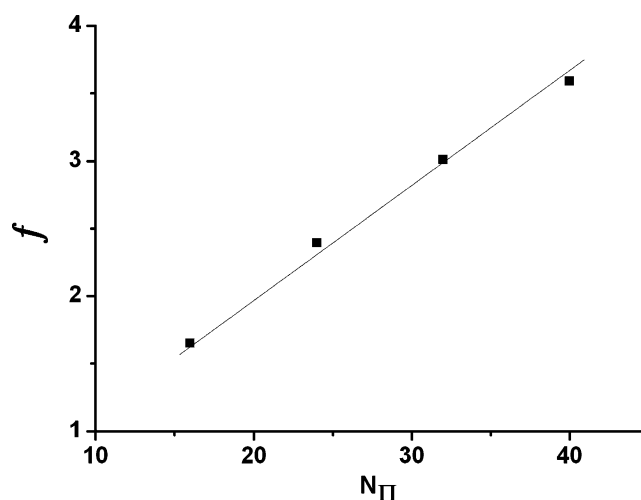


Fig. 9 The dependence of f on N_π

absolute value of β_0 increases as the π -conjugated ring gets larger. β_0 almost doubles each time another $C[3T_DA]$ unit is added to the ring.

In order to elucidate the origin of the static second-order nonlinear optical susceptibilities of these molecules, we scrutinized several main excited states. For example, for the $C[3T_DA]_2$ molecule, the fourth and fifth excited states make considerable contributions to β_0 (25.27% and 37.74%, respectively), so the fifth excited state (S_5) was selected for further study in our work. Similarly, the transitions from S_0 to S_3 in $C[3T_DA]_3$, $C[3T_DA]_4$ and $C[3T_DA]_5$ were picked for the following discussion. The important excited states were HOMO→LUMO+1 and HOMO→LUMO+2 transitions.

According to the two-level approximation formula $\beta_0 = 3e^2\hbar^2f\Delta\mu/2mW^3 = Bf\Delta\mu/W^3$, three factors impact on β_0 : $\Delta\mu$ (the difference in the dipole moments between the excited state and the ground state), W (the transition

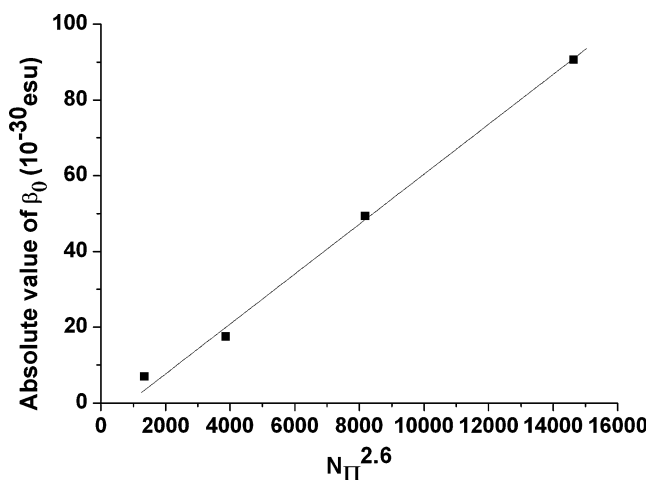


Fig. 10 The dependence of absolute β_0 on $N_\pi^{2.6}$

Table 4 Parameters for C[3T_DA]_n and C[3T_DA]_n-L calculated with the FF method

FF (a.u.)	β_{xxx}	β_{xxy}	β_{xyy}	β_{yyy}	β_{xxz}	β_{yyz}	β_{xzz}	β_{yzz}	β_{zzz}
C[3T_DA] ₂	3.841	-20.042	-40.930	-52.206	-17.705	47.557	46.576	-113.516	-58.488
C[3T_DA] ₃	787.761	725.718	-83.153	62.120	2199.095	-55.760	-696.710	-357.732	-2256.394
C[3T_DA] ₄	88.488	-327.465	-39.171	-194.714	-115.423	160.246	19.402	-231.413	-314.101
C[3T_DA] ₅	407.284	362.564	9.946	383.246	106.892	8.268	132.149	354.174	274.925
C[3T_DA] ₂ -L	-404.530	-3.395	-31.878	0.001	166.135	88.321	466.510	-0.419	-355.119
C[3T_DA] ₃ -L	400.986	481.726	-26.971	-293.100	-373.852	-42.193	28.094	60.615	64.423
C[3T_DA] ₄ -L	209.238	-201.578	149.676	195.022	1483.069	-16.580	549.931	31.275	573.151
C[3T_DA] ₅ -L	1785.251	-695.658	245.539	-795.961	396.072	-317.166	-604.067	-659.410	-849.579

FF (a.u.)	μ_x	μ_y	μ_z	β_x	β_y	β_z	$\beta_0 \times 10^{-30}$ (esu)
C[3T_DA] ₂	-0.018	0.315	0.050	9.487	-185.764	-28.636	-1.62
C[3T_DA] ₃	-0.014	-0.478	0.120	7.898	430.106	-113.050	-3.84
C[3T_DA] ₄	-0.069	0.773	0.273	68.718	-753.592	-269.678	-6.94
C[3T_DA] ₅	0.001	-1.261	0.001	549.379	1099.984	390.085	-9.50
C[3T_DA] ₂ -L	-0.702	0.000	1.777	30.103	-3.813	-100.663	-0.90
C[3T_DA] ₃ -L	-0.154	0.976	-0.388	402.111	249.241	-351.622	2.59
C[3T_DA] ₄ -L	-0.804	-0.577	1.804	908.843	24.720	2039.640	12.32
C[3T_DA] ₅ -L	0.925	-1.624	-1.098	1426.723	-2151.029	-770.673	22.56

energy), and f (oscillator strength). It can be seen from Table 3 that the dipole moments of the ground state (μ_0) and the excited state (μ_k) increase gradually from C[3T_DA]₂ to C[3T_DA]₅ as N_π increases, but the absolute value of $\Delta\mu$ only increases slowly. The relation $\Delta\mu \propto N_\pi$ is shown in Fig. 6.

From Table 3, it is apparent that W decreases with increasing the number of units, so the denominator decreases as the cube of W , which causes the absolute value of β_0 to increase significantly. It is known from [46] that W can be expressed by the following formula: $W = a + b \sin(\pi/(2N+1))$, $N = N_\pi + t$, where t is defined as the end group effect index. Here, t is zero because the molecules studied in this work are annular. The dependence of W on $\sin[\pi/(2N+1)]$ is plotted in Fig. 7, which shows a linear relationship. The variation of $\lg W$ with $\lg N_\pi$ is shown in Fig. 8; it is a line with a slope of -0.2 ; thus, the equation $W = AN_\pi^{-0.2}$ is

obtained, and the intercept distance A is about $44,500 \text{ cm}^{-1}$. The dependence of f on N_π is also linear (as shown in Fig. 9).

In a word, N_π increases with the number of units, $\Delta\mu \propto N_\pi$, $f \propto N_\pi$, and $W \propto N_\pi^{-0.2}$, so that $\beta_0 \propto N_\pi^{2.6}$. The relationship between the absolute β_0 and $N_\pi^{2.6}$ is illustrated in Fig. 10, which suggests that enlarging the ring causes N_π and the conjugated range to increase; as a result, $\Delta\mu$ and f increase while W decreases, which causes the absolute value of β_0 to increase. Moreover, this clearly indicates that W is the dominant factor in the increase of β_0 .

Compared to the traditional thiophene derivatives reported previously [47], the calculated absolute β_0 values of the annular C[3T_DA]_n species are larger. However, if we consider the linear structures (C[3T_DA]_n-L, $n=2-5$) that are obtained from the corresponding four annular species (C[3T_DA]_n), the results in Table 3 suggest that the

Table 5 BLA values and related parameters for C[3T_DA]_n and C[3T_DA]_n-L

Ring molecule	Sum of S bond lengths	Average S bond length	Sum of D bond lengths	Average D bond length	BLA value
C[3T_DA] ₂	22.6433	1.4152	21.4248	1.3390	0.0762
C[3T_DA] ₃	33.8941	1.4122	32.1271	1.3386	0.0736
C[3T_DA] ₄	45.1790	1.4118	42.8267	1.3383	0.0735
C[3T_DA] ₅	56.5052	1.4126	53.5140	1.3378	0.0748
C[3T_DA] ₂ -L	21.2707	1.4180	21.5333	1.3458	0.0722
C[3T_DA] ₃ -L	32.5641	1.4158	32.3198	1.3466	0.0692
C[3T_DA] ₄ -L	43.8302	1.4139	42.8142	1.3379	0.0760
C[3T_DA] ₅ -L	55.4088	1.4207	53.8117	1.3453	0.0754

absolute value of β_0 increases as N_π and f increase and W decreases, while the slightly decrease in $\Delta\mu$ values does not influence the increasing trend of β_0 . Comparing the three factors for the two types of molecules, both f and $\Delta\mu$ for the linear molecules are bigger than those for the annular molecules, while their W values are smaller. This means that the β_0 values of the linear molecules are larger than those seen for the ring molecules. The reason for this may be the enhanced π -conjugation of the first two linear molecules (i.e., $n=2, 3$), and the larger polarities of the latter two ($n=4, 5$).

Aside from the ZINDO-SOS method, another popular method, the finite field (FF) method, was also used to calculate the static second-order nonlinear optical susceptibility. The calculated β_0 values are listed in Table 4, which also shows that the absolute value of β_0 increases with N_π , and that the β_0 values of the linear molecules are larger than those of the ring molecules. If we compare the values of β_0 calculated by the two methods, they are different but they show a similar trend. This reflects the fact that the absolute value of β_0 gradually increases with the number of units, despite the different values of β_0 obtained with the two methods. This result indicates new ways to achieve high static second-order nonlinear optical susceptibility: by enlarging the ring size or the length of the conjugated chain.

It is well known that the second-order nonlinear optical coefficient (β) is closely related to the bond length alternation (BLA) [48]. As seen from the structures of the studied molecules (Fig. 1), the whole outer ring skeleton of each molecule appears to be an alternating arrangement of single and double bonds. In order to study the relationship between the second-order nonlinear optical coefficient (β) and BLA, the sum and the average of the single carbon–carbon bond (“S bond”) lengths and the double or triple carbon–carbon bond (“D bond”) lengths in the π -conjugated rings were calculated, and then their BLAs were obtained (Table 5). It can be seen from Table 5 that the S bond length gets closer to that of the D bond as the ring enlarges, while the BLA decreases as the number of trithiophene and diacetylene units increase (except in the case of C[3T_DA]₅), indicating that the conjugation of the whole ring increases with ring size. The middle thiophene of C[3T_DA]₅ shows serious distortion, which may influence the π -conjugation and yield an increased BLA. To some extent, the decrease in BLA with the number of units may also decrease the band gap. The BLAs of the linear molecules (C[3T_DA]_{*n*}-L) were also calculated for the sake of comparison. As shown in Table 5, the BLA of C[3T_DA]₃-L is smaller than that of C[3T_DA]₂-L, and both of those values are smaller than the BLAs of C[3T_DA]₂ and C[3T_DA]₃, which clearly shows that those two linear molecules enhanced π -conjugation. The BLAs of the latter two linear molecules, i.e., C[3T_DA]₄-L and C[3T_DA]₅-

L, are larger than those of C[3T_DA]₂-L and C[3T_DA]₃-L and larger than those for the corresponding ring structures, due to their increased polarities. Thus, the conjugation of the ring improves as the number of units increases, but the conjugation of the linear skeleton does not. In a word, the conjugation across the whole skeleton plays an important role in the electronic structure and the static second-order nonlinear optical coefficient of each molecule.

Conclusions

In this study, the geometric structures, UV-Vis spectra and electronic structures of a series of macrocyclic thiophene derivatives (C[3T_DA]_{*n*}) with C_n symmetry ($n=2-5$) were studied theoretically. The results show that ΔE_{H-L+1} and ΔE_{H-L+2} decrease as the number of C[3T_DA] units increases, and the UV-Vis spectrum redshifts, but ΔE_{H-L} increases. The static second-order nonlinear optical coefficient (β_0) of the series of macrocycles C[3T_DA]_{*n*} with C_n symmetry, and those of their corresponding linear structures C[3T_DA]_{*n*}-L, were evaluated by the ZINDO-SOS and FF methods. The calculated absolute value of β_0 increases with the number of units. Both W and f have a large effect on the value of β_0 . The absolute β_0 values of the macrocyclic thiophene derivatives are smaller than those of their corresponding linear molecules due to their smaller polarities, but they are larger than traditional thiophene derivatives because of their enhanced π -conjugation. This provides a useful guideline for the design of efficient macrocyclic derivatives with large static second-order nonlinear optical susceptibilities. Overall, this series of compounds are candidates for second-order nonlinear optical materials. Obviously, the theoretical study performed in this paper is just the start of research into the NLO properties of compounds in this family.

Acknowledgments This work was supported by the Natural Science Foundation of China (nos. 20973078 and 20673045), by special funding given to basic scientific research projects for Central Colleges, by the Open Project of the State Key Laboratory for Supramolecular Structure and Material of Jilin University (SKLSSM200716), and by the Graduate Innovation Fund of Jilin University (no. 20111031).

References

1. Jen AKY, Cai Y, Bedworth PV, Marder SR (1997) *Adv Mater* 9:132–135
2. Wu X, Wu J, Liu Y, Jen AKY (1999) *J Am Chem Soc* 121:472–473
3. Sun SS, Zhang C, Dalton LR, Garner SM, Chen A, Steier WH (1996) *Chem Mater* 8:2539–2541

4. Jen AKY, Liu Y, Zheng L, Liu S, Drost KJ, Zhang Y, Dalton LR (1999) *Adv Mater* 11:452–455
5. Cai C, Liakatas I, Wong MS, Bösch M, Bosshard C, Günter P, Concilio S, Tirelli N, Suter UW (1999) *Org Lett* 1:1847–1849
6. Morley JO (1991) *J Chem Soc Faraday Trans* 87:3009–3013
7. Zhao HB, Qiu YQ, Liu CG, Sun SL, Liu Y, Wang RS (2010) *J Org Chem* 69:2251–2257
8. Kim HM, Cho BR (2009) *J Mater Chem* 19:7402–7409
9. Carella A, Castaldo A, Centore R, Fort A, Sirigu A, Tuzi A (2002) *J Chem Soc Perkin Trans* 2:1791–1795
10. Batista RMF, Costa SPG, Belsley M, Raposo MMM (2007) *Tetrahedron* 63:9842–9849
11. Krawczyk P (2010) *J Mol Model* 16:659–668
12. Kimur T, Kato M, Duan XM, Matsuda H, Fukuda T, Okada S, Nakanishi H (2000) *Macromol Chem Phys* 201:178–183
13. Li J, Feng JK, Sun CC (1994) *J Phys Chem* 98:8636–8640
14. Bartkowiak W, Misiaszek T (2000) *Chem Phys* 261:353–357
15. Bhaskar A, Ramakrishna G, Hagedorn K, Varnavski O, Osteritz EM, Bauerle P, Goodson T III (2007) *J Phys Chem B* 111:946–954
16. Fuhrmann GL (2006) Synthesis and characterization of oligothiophene-based fully π -conjugated macrocycles. University of Ulm, Germany
17. Fuhrmann G, Bäuerle P, Casadoa J, Zgierski MZ, Navarrete JTL (2006) *J Chem Phys* 125:044518
18. Zhao Y, Truhlar DG (2004) *J Phys Chem A* 108:6908–6918
19. Zhao Y, Truhlar DG (2005) *J Chem Theory Comput* 1:415–432
20. Perdew JP (1991) In: Ziesche P, Eschig H (eds) *Electronic structure of solids '91*. Akademie Verlag, Berlin, p 11
21. Becke AD (1996) *J Chem Phys* 104:1040–1046
22. Ridley J, Zerner MC (1973) *Theor Chim Acta* 32:111–134
23. Krätschmer W, Lamb LD, Fostiropoulos K, Human DR (1990) *Nature* 347:354–358
24. Lyoda M, Sultana F, Sasaki S, Yoshida M (1994) *J Chem Soc Chem Commun* 1929–1930
25. Orr BJ, Ward TF (1971) *Mol Phys* 20:513–526
26. Kanis DR, Ratner MA, Marks TJ (1994) *Chem Rev* 94:195–242
27. Bishop DM (1994) *Adv Quantum Chem* 25:1–3
28. Ulman A, Willand CS, Kohler W, Tobello DR, Williams DJ, Handley L (1990) *J Am Chem Soc* 112:7083–7090
29. Fu W, Feng JK, Pan GB, Zhang X (2001) *Theor Chem Acc* 106:241–250
30. Feng JK, Gao XL, Sun CC (1992) *Chin Sci Bull* 37:1441–1445
31. Gao XL, Feng JK, Sun CC (1992) *Int J Quantum Chem* 42:1747–1758
32. Li J, Feng JK, Sun CC (1994) *Int J Quantum Chem* 52:673–680
33. Ren AM, Feng JK, Zhao XX, Liu CL, Su ZM (2001) *Chem J Chin Univ* 22:1197–1200
34. Fu W, Feng JK, Pan GB (2001) *J Mol Struct THEOCHEM* 545:157–165
35. LeCours SM, Guan HW, DiMugno SG, Wang CH, Therien MJ (1996) *J Am Chem Soc* 118:1497–1503
36. Kanis DR, Ratner MA, Marks TJ (1992) *J Am Chem Soc* 114:10338–10357
37. Oudar JL, Chemla DS (1977) *J Chem Phys* 66:2664–2668
38. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (2003) *Gaussian 03*, revision D02. Gaussian Inc., Pittsburgh
39. Thanthiriwate KS, Silva KMND (2002) *J Mol Struct THEOCHEM* 617:169–175
40. Liyanage PS, Silva RMD, Silva KMND (2003) *J Mol Struct THEOCHEM* 639:195–201
41. Kurtz HA, Stewart JJP, Dieter K (1990) *J Comput Chem* 11:82–87
42. Huang S, Ren AM, Li Z, Zhao Y, Min CG (2010) *Chem J Chin Univ* 3:553–558
43. Fuhrmann G, Debaerdemaeker T, Bäuerle P (2003) *Chem Commun* 948–949
44. Kromer J, Rios-Carreras I, Fuhrmann G, Musch C, Wunderlin M, Debaerdemaeker T, Mena-Osteritz E, Bauerle (2000) *Angew Chem Int Ed* 39:3481–3486
45. Feng JK (2005) *Acta Chim Sinica* 63:1245–1256
46. Xu GX, Li LM (1980) *Sci China* 2:136–151
47. Chou SSP, Sun DJ, Huang JY, Yang PK, Linc HC (1996) *Tetrahedron Lett* 37:7279–7282
48. Marder SR, Perry JW, Bourhill G, Gorman CB, Tiemann BG, Mansour K (1993) *Science* 261:186–189

Architecture, electronic structure and stability of TM@Ge(n) (TM = Ti, Zr and Hf; n = 1-20) clusters: a density functional modeling

Manish Kumar · Nilanjana Bhattacharyya ·
Debashis Bandyopadhyay

Received: 23 February 2011 / Accepted: 9 May 2011 / Published online: 28 May 2011
© Springer-Verlag 2011

Abstract The present study reports the geometry, electronic structure and properties of neutral and anionic transition metal (TM = Ti, Zr and Hf) doped germanium clusters containing 1 to 20 germanium atoms within the framework of linear combination of atomic orbitals density functional theory under spin polarized generalized gradient approximation. Different parameters, like, binding energy (BE), embedding energy (EE), energy gap between the highest occupied and lowest unoccupied molecular orbitals (HOMO-LUMO), ionization energy (IP), electron affinity (EA), chemical potential etc. of the energetically stable clusters (ground state cluster) in each size are calculated. From the variation of these parameters with the size of the clusters the most stable cluster within the range of calculation is identified. It is found that the clusters having 20 valence electrons turn out to be relatively more stable in both the neutral and the anionic series. The sharp drop in IP as the valence electron count increases from 20 to 21 in neutral cluster is in agreement with predictions of shell models. To study the vibrational nature of the clusters, IR and Raman spectrum of some selected TM@Ge_n (n = 15, 16, 17) clusters are also calculated and compared. In the end, relevance of calculated results to the design of Ge-based super-atoms is discussed.

Keywords Binding energy · Clusters and nanoclusters · Density functional theory · Electron affinity · Embedding energy · Ionization potential · IR and Raman

M. Kumar · N. Bhattacharyya · D. Bandyopadhyay (✉)
Physics Department, Birla Institute of Technology and Science,
Pilani, Rajasthan 333031, India
e-mail: Debashis.bandy@gmail.com

D. Bandyopadhyay
e-mail: bandy@bits-pilani.ac.in

Introduction

In the last few decades study of electronic structures and properties of silicon and germanium based semiconductor nanoclusters has been an extremely active area of research due to its importance in nanoscience and nanotechnology [1–9]. Theoretical modeling of semiconductor nanoclusters with modified new geometries show possibility of using them as nano-devices in opto-electronics, tunable lasers, sensors, etc. The results of the theoretical modeling also show that it is possible to integrate wide area of photonics with electronics by designing miniature devices using pure and hybrid semiconductor nanoclusters [10, 11]. Therefore, modeling of pure and hybrid semiconductor nanoclusters is a very interesting and important field in theoretical and computational research with supporting experimental results. It is always challenging to produce experimentally or model theoretically the stable and nonreactive nanoclusters of different compositions of interest from the application point of view. It is known that pure semiconductor nanoclusters are chemically reactive, but encapsulation of transition metal atom/atoms by pure semiconductor clusters enhances the stability of the clusters and simultaneously exhibits much novel behavior. The increase in stability of the semiconductor clusters when doped with transition metal atom/atoms is due to the fact that the transition metal atom can absorb the dangling bonds present on the pure semiconductor cage or cluster surface [12–17]. An important experimental contribution on such nanoclusters was made by Beck [18, 19] by using laser vaporization supersonic expansion technique. It showed that addition of the transition metals like, Cr, Mo W etc. in silicon clusters enhanced the stability of the doped clusters compared to the same size pure silicon semiconductor clusters by subjecting them to photo-fragmentation. Hiura

et al. [7] reported the formation of a series of stable silicon cages with transition-metal atoms Hf, Ta, W, Re, Ir etc. O'hara et al. [20] studied the geometric and electronic structures of the negatively charged Tb doped silicon clusters with photoelectron spectroscopy and chemical-probe method, and proved that Tb atom always remains encapsulated inside the silicon clusters made of ten silicon atoms. Recently, Bandyopadhyay [15, 16] reported an extensive study of the electronic structure, growth behavior, different physical and chemical properties of the pure and transition metal doped silicon clusters (TM@Si_n, TM = Ti, Zr and Hf, n=9–20). It is found that the metal-doped fullerene like Si₁₆TM (TM = Ti, Zr and Hf) clusters are chemically the most stable species in the whole series of study. Compare to silicon nanoclusters there is not much report on the investigation of germanium nanoclusters theoretically as well as experimentally. Some of the recent investigations on pure and halogen-doped germanium clusters are mainly focused on optimized geometries, binding energies, ionization potential, electron affinities etc. [2, 7, 18–21]. A few theoretical and experimental contributions have been made by different groups on the endohedral doping of transition metal (TM) elements in pure and hydrogenated Ge cages [22–25]. Even after several studies of pure and hybrid silicon and germanium nanoclusters, there is no clear scientific way to explain the cause of the stabilities of nanoclusters of a particular size in a system. From famous “electron counting rule” point of view it is not clear whether their relative stability obeys 18-electron counting rule (also known as the octet rule) or 20-electron counting rule [26]. Following 18-electron rule Si₁₂Cr and Si₁₂W should be the most stable clusters in the 3 d and 5 d TM doped Si_n series respectively. Recent investigations showed the electron counting rule sometimes a valid explanation for stability, but not always [27–29]. It is well known that the stability depends upon the nature of transition metal atoms. In addition, one needs to follow the free-electron gas theory and Wigner-Witmer (WW) spin conservation rule [30] while calculating embedding energies (EE) to explain the stability of the cluster. However, the free-electron gas picture is not valid in every case because it is found that the anionic Si₁₂Mn and Si₁₂Co show maximum value in embedding energy curve whereas they are 20 and 22-electron clusters respectively according to electron counting rule [29]. On the other hand, experiments have supported the validity of these electron-counting rules in some cases. Koyasu et al. [31] studied the electronic and geometrical structures of Si₁₆TM (TM = Sc, Ti, and V) clusters using mass spectrometry and anion photoelectron spectroscopy. They found that neutral Si₁₆Ti, being a 20-electron cluster, had a closed-shell electron configuration with a large HOMO-LUMO gap. Recently Bandyopadhyay

[15, 16] also found similar behavior in density functional studies of TM@Si_n (TM = Ti, Zr and Hf) system.

With this background work, the present study makes an effort to explain the enhanced stability of TM@Ge₁₆ (TM = Ti, Zr and Hf) clusters by calculating different parameters of the ground state cluster in each size within the size range of 1 to 20 germanium atoms using density functional theory (DFT). Although Ti-doped Ge_n clusters have been theoretically studied before [32], this study was limited to a particular size range. The main focus of the present study is to explain the relative stability of these clusters in neutral and cationic state in the range from n =1–20 in view of electron counting rules and also to study their chemical and vibrational properties.

Computational methods

In the present report the theoretical calculations are performed within the framework of linear combination of atomic orbitals density functional theory under spin polarized generalized gradient approximation. The exchange-correlation potential contributions are incorporated in the calculation by using generalized gradient functional (GGA) proposed by Perdew-Wang (B3PW91) [33–36]. Molecular orbitals are expressed as linear combination of atom-centered basis functions for which the standard Gaussian basis (LanL2DZ) associated with effective core potential (ECP) is used on all atoms. This basis set can reduce the difficulties in two-electron integrals caused by the transition metal atoms [37–40]. All geometry optimization calculations are performed under no external symmetry constrictions. It is always possible that during optimization the cluster can get trapped in local minima of the potential energy surface. To avoid it during optimization in a particular size, several geometrical initial guess configuration at different spin states (sometimes varied from singlet to septet) were used to search for the ground state and higher energy isomers. In case when total optimization energy decreases with increasing spin, increasing higher spin states is considered until energy minima with respect to spin is reached. For small clusters, considered in this work, such an approach provides a fairly extensive search of the potential energy minima. In order to check the validity of the applied methodologies, a trial calculation is carried out on Ge-Ge dimer. Calculated Ge-Ge bond length is 2.54 Å which is within the range of the values obtained theoretically as well as experimentally reported by Nagendran et al. [41]. The optimized electronic structure was obtained by solving the Kohn–Sham equations self-consistently [42] where default optimization criteria of Gaussssian-03 [43] are followed. Initial geometries of the clusters are constructed on the basis of the reported structures [2, 3] and from

intuitions. Equilibrium structure in a particular size is obtained by varying the geometry starting from high to low symmetric initial guess structures. For each stationary point of a cluster, the reliability of the clusters is reassured by calculating the frequency of harmonic vibration (IR and Raman). If any imaginary frequency is found, a relaxation along that vibrational mode is carried out until the true local minimum is obtained. Broadening is applied on the calculated frequencies to plot the IR and Raman response of the clusters, which is a default criterion of Gauss View 03 program package. In each size, identical sets of initial geometries are used for Ti, Zr and Hf doped germanium clusters to see their growth behavior. Geometry optimizations were carried out with a convergence limit of 10^{-7} Hartree on the total electronic energy. The optimized electronic structure for each cluster is obtained from the *Z*-matrices in the program output, where as electronic properties were calculated from the SCF total electronic energy and the orbital energy values. All theoretical calculations are performed with GAUSSIAN-03 program package [43].

Results and discussion

In our previous report [11] a detailed study of the growth behavior of pure germanium cluster is presented. Therefore, in the report the growth behavior of different transition metal (Ti, Zr and Hf) doped germanium clusters will be discussed.

Growth of hybrid Ge_nTM nanoclusters

Since the growth evolution pattern of Ti, Zr and Hf are similar, therefore, in the following section the growth behavior of Ti@Ge_n will be presented with related discussion of Zr and Hf doped germanium clusters. Calculated optimized geometries within the size range of $n=1-20$ are shown in Fig. 1 with the relative energy of the ground state Ti@Ge_n clusters in each size. In the same figure the ground state geometries of Zr and Hf doped clusters for $n=9-20$ size range are also presented. A number of isomers are calculated in each size. But only selected isomers those are close to the ground state structures are presented in Fig. 1. The present report is mainly based on the Ti doped germanium cluster and hence it is necessary to mention the basic properties of titanium. It is well known that titanium is paramagnetic material with $\text{Ar}3d^24s^2$ electronic configuration and electro-negativity 1.91 in Pauling scale. When germanium cluster doped with titanium, in most of the case sp^3d^2 hybridization with the germanium atoms in the clusters takes place.

The first structure of Ge_nTM (TM = Ti, Zr and Hf) series is a dimer optimized for singlet, triplet, quintet and septet

spin states. It is found that Ge-Ti dimer in quintet spin state is the optimized ground state (GS). The next ground state optimized structure for $n=2$ is in triplet spin state for all TM doped clusters. At this size, only four possible geometries can be constructed and all four different structures are optimized. Among these four, the triangular structure $\text{Ge}_2(\text{A})\text{TM}$ is the optimized ground state structure. The second structure is a bent structure (Ge-TM-Ge) with an internal angle 44.67° for Ti@Ge_2 . Other two structures are linear chain structures with the metal atoms are positioned at the end (Ge-Ge-TM) or at the middle of the chain (Ge-TM-Ge). In Ge_3TM , three different optimized geometries (two rhombi and a pyramid) are optimized. In the planer rhombi structures position of TM metal atoms are different as shown in Fig. 1. The rhombus like structure $\text{Ge}_3(\text{A})\text{TM}$, in triplet spins is found as ground state.

In Ge_4TM family, five different isomers are optimized with a $\text{Ge}_4(\text{A})\text{TM}$ in triplet spin state as GS structure (Fig. 1). This structure can be obtained by capping a TM atom on a Ge_4 (bent rhombus) structure. Other optimized isomers are mainly planar structures with transition metal atom positioned at different sites. These structures are ‘planar pentagon’ and ‘rhombus with a tail’ like structures in different forms as shown in Fig. 1. Three different structures are obtained in Ge_5TM size (Fig. 1) with small difference in their optimized energy. The ground state structure in triplet spin state looks like a bi-capped quadrilateral where one capped atom is Ge and the other one is TM. Change in position of TM atom in the cluster can give other isomers as shown in Fig. 1. Three different optimized geometries in Ge_6TM series are shown in Fig. 1. These structures are based on the ground state structure of $\text{Ge}_6(\text{A})$ and $\text{Ge}_7(\text{A})$ as reported by Bandyopadhyay and Sen [11]. The optimized triplet ground state structure $\text{Ge}_6(\text{A})\text{TM}$ is obtained by adding one TM atom on the side triangular plane of $\text{Ge}_6(\text{A})$ ground state structure or by replacing the capped germanium atom by TM and adding one extra germanium atom on the bi-capped pyramidal plane of $\text{Ge}_6(\text{A})$ structure. Other structures shown in Fig. 1 can be obtained by replacing one germanium atom in $\text{Ge}_7(\text{A})$ cluster [11] from different positions. Addition of one TM atom with one of the side planes of $\text{Ge}_7(\text{A})$ pure structure gives optimized ground state structure in triplet spin state $\text{Ge}_7(\text{B})\text{TM}$ as shown in Fig. 1. With increasing the number of germanium atoms in the cluster, the tendency to encapsulate TM atom in stable structures increase. In the ground state $\text{Ge}_8(\text{C})\text{Ti}$ structure, the Ti atom is outside the germanium cluster. It is possible to optimized a structure with endohedrally capped TM although it is not the GS structure in this size. Following the reported geometries available in literatures [2, 3] and also some more new geometry constructed from intuition, with an aim to encap-

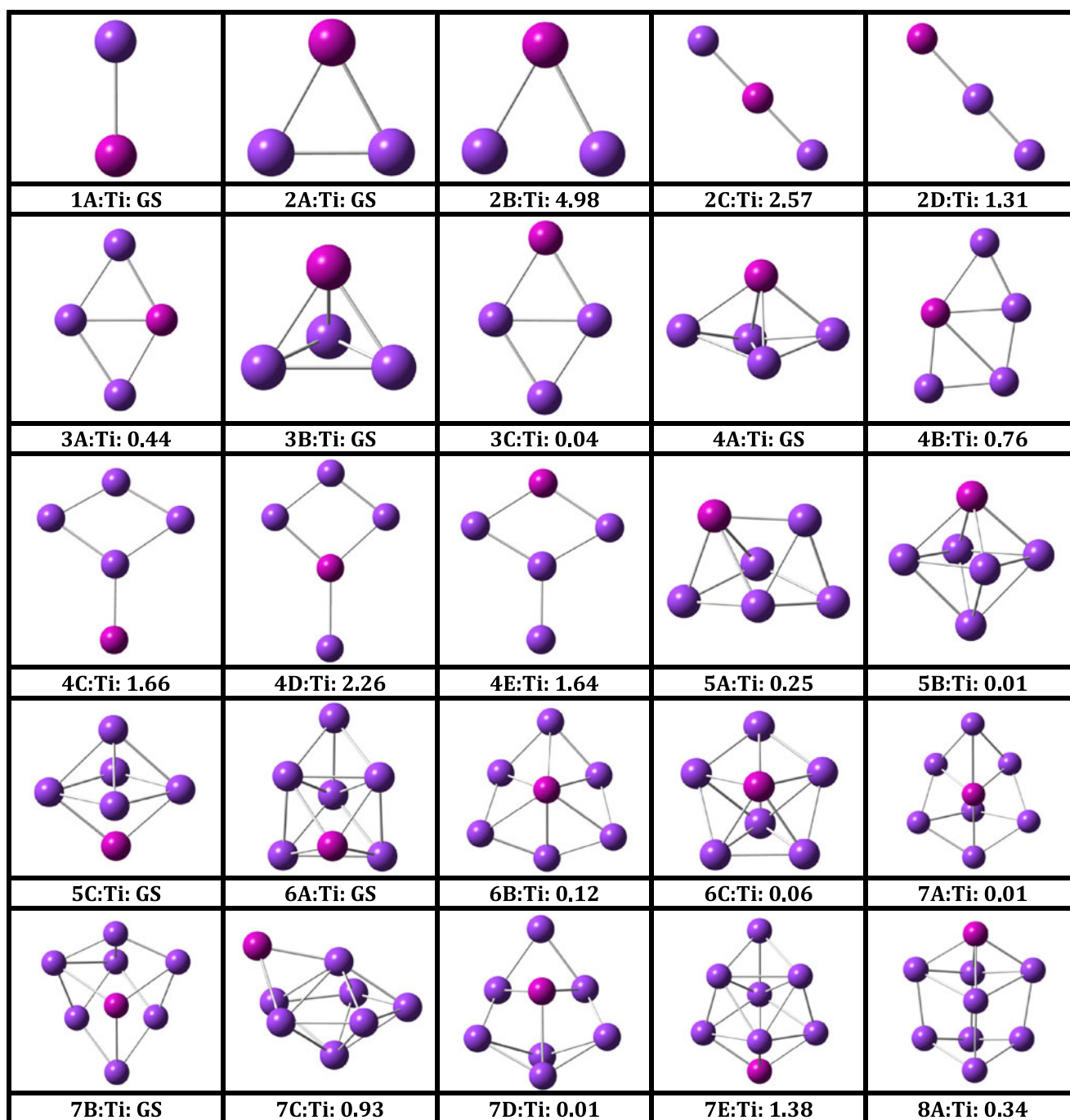


Fig. 1 Optimized structures of pure Ge_nTi and some selected Ge_nZr and Ge_nHf clusters for $n=1$ to 20. The dip purple colored solid spheres in the clusters represent Ge atoms. The numeric given next to

the cluster name represents the relative energies of the optimized geometries in eV with respect to the ground state cluster of same size

simulate the titanium atom endohedrally, three different geometries are taken as initial guess structures to search for the optimized ground state in $\text{TM}@Ge_n$ series. The structures $\text{Ge}_9(\text{A})\text{TM}$ and $\text{Ge}_9(\text{C})\text{TM}$ as shown in Fig. 1 absorb transition metal atom endohedrally in optimized state. The third structure $\text{Ge}_9(\text{B})\text{TM}$ is the ground state and it absorbs the transition metal atom exohedrally on the Ge_9 surface. The

optimized ground state isomer $\text{Ge}_9(\text{B})\text{Zr}$ and $\text{Ge}(\text{B})\text{Hf}$ also absorb the TM atoms exohedrally on the Ge_9 cluster surface. Therefore up to this size the endohedrally doped structures are energetically not favorable.

In the next series, Ge_{10}Ti , a number of optimized structures are found; out of those structures only four different structures have been reported here. The first

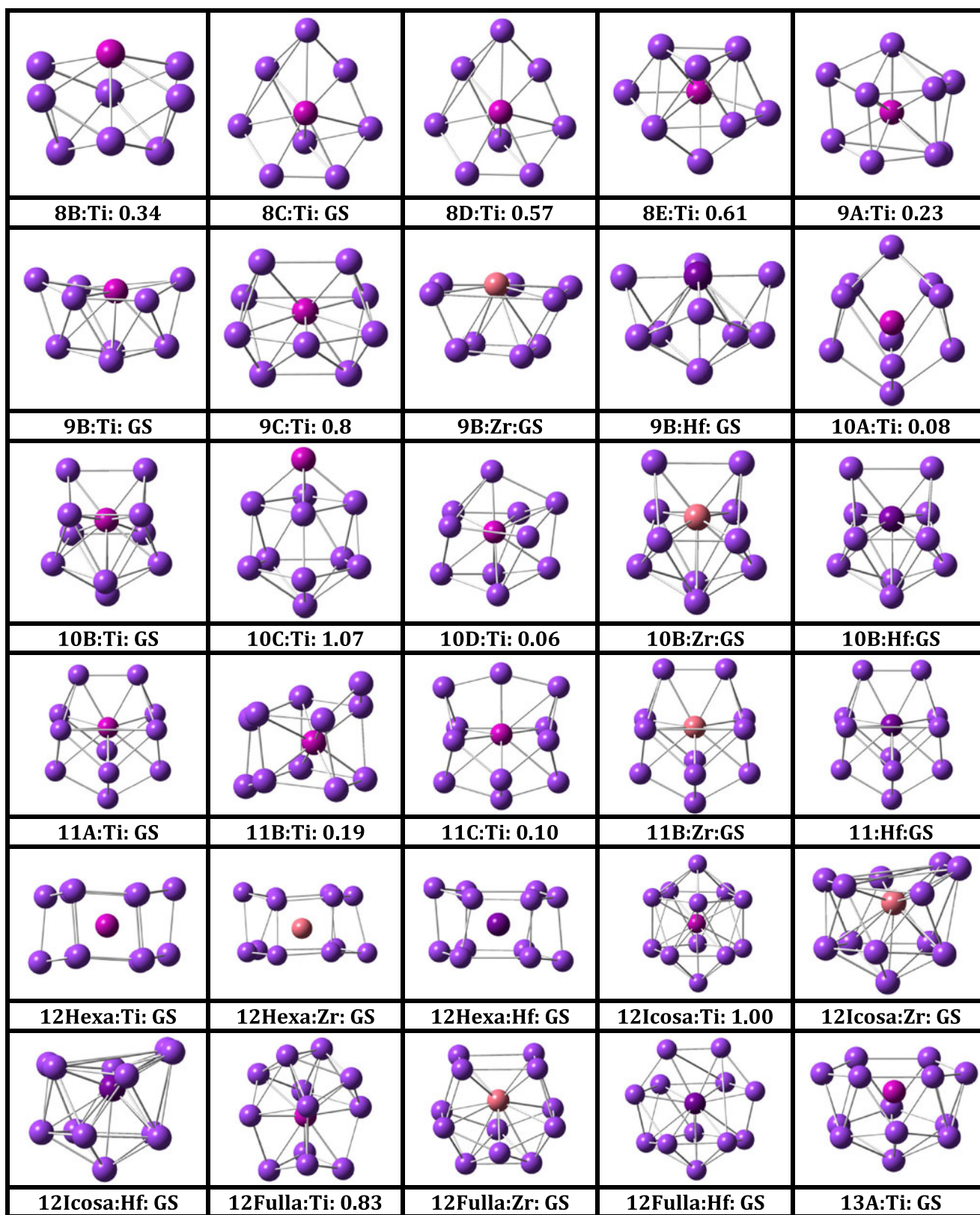


Fig. 1 (continued)

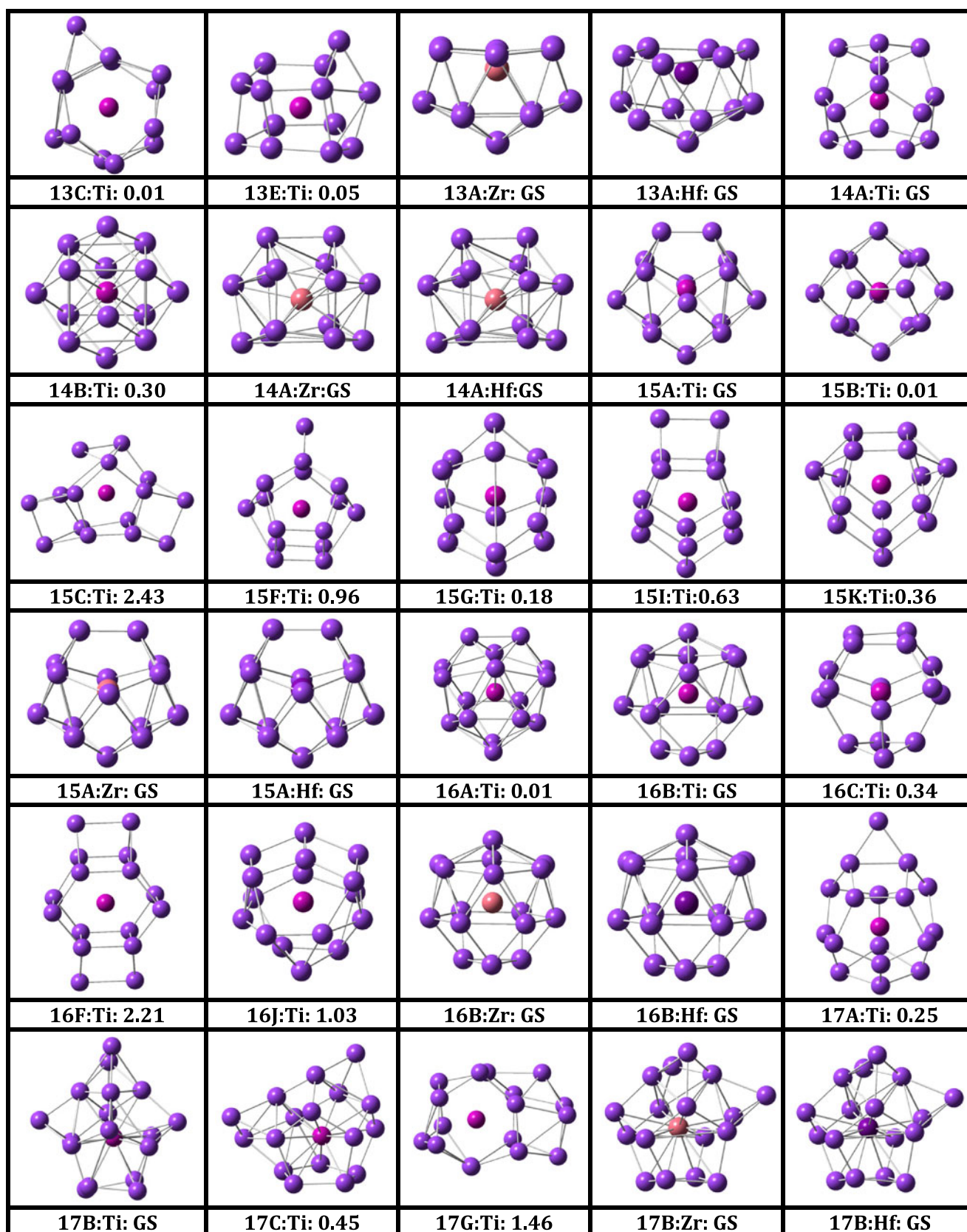


Fig. 1 (continued)

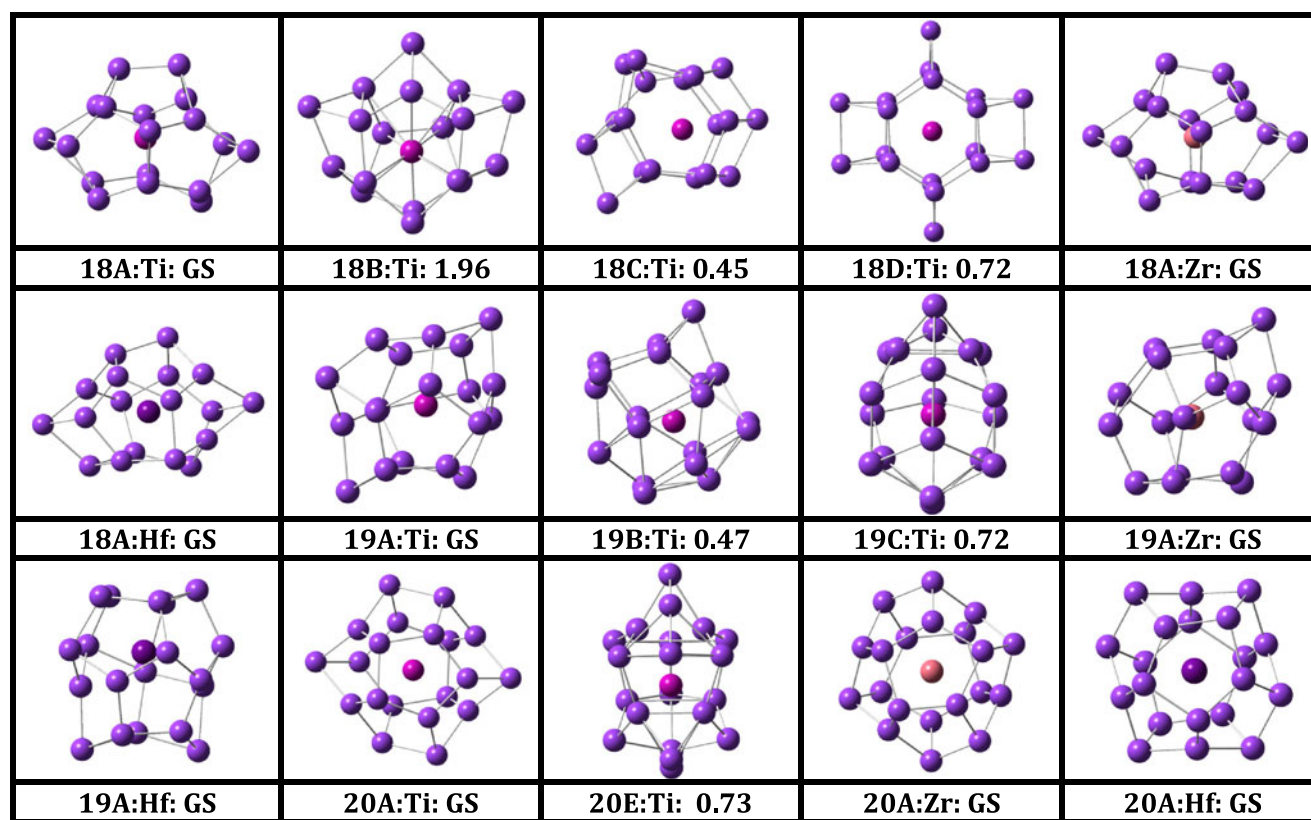


Fig. 1 (continued)

optimized structure $\text{Ge}_{10}(\text{A})\text{Ti}$ is a combination of pentagonal and rhombus planes capped by the germanium atom. The second structure $\text{Ge}_{10}(\text{B})\text{Ti}$, is the first in this series that can give GS geometry by absorbing Ti atom endohedrally as shown in Fig. 1. The third and the fourth structures absorb Ti atom exohedrally and endohedrally respectively. For $n=11$ we obtained three endohedrally absorbed TM doped structures. The structure $\text{Ge}_{11}(\text{A})\text{TM}$ is the ground state for all Ti, Zr and Hf as shown in Fig. 1. Out of five optimized geometries, three important optimized geometries, known as hexagonal, fullerene and icosahedral structures in Ge_{12}TM series are presented in Fig. 1. Among these three structures in different spin states, the relaxed hexagonal singlet spin state structure is the ground state. In some transition metal doped Ge_{12} fullerene-like structures (combination of four pentagons and four rhombi) show ground state. It is known that this kind of pentagon and rhombi combination always gives better stability [14]. In this context it is important to mention that transition metal-doped fullerene-like structure is not always the ground state isomer. It depends upon the nature of transition metal doping [39]. As example the Cu doped fullerene structure is not a ground state structure for $n=12$ size. According to previous investigations on the transition metal (Ti, Zr and Hf) encapsulated silicon clusters, the metal-encapsulated hexagonal prism is proven to be the lowest-energy structure [15, 16]. In another study the

optimized Ge_{12}Zn icosahedral isomer is lower in energy than the hexagonal prism structure [44, 45]. Therefore, it is clear that in $n=12$ size, the ground state structure is strongly depends on the transition metal doping. Careful observation show that the ground state structures from $\text{Ge}_9(\text{A})\text{Ti}$ to $\text{Ge}_{11}(\text{A})\text{Ti}$ are all of similar kinds and follow a common trend in their growth behavior. Several initial guess geometries are optimized in Ge_{13}TM series. Out of them three optimized structures including the ground state $\text{Ge}_{13}(\text{A})\text{TM}$ geometry are presented in Fig. 1. The other geometries are very similar in structure to the GS state geometry with a minor difference in energy. Therefore in the present report those geometries are not presented. The ground state $\text{Ge}_{13}(\text{A})\text{TM}$ structure is hexagonal capped bowl kind of structure with transition metal atom inside. It is to be noted that the binding energy presented in Fig. 2 show clear two trend of variation of binding energy with the size of the cluster for pure and doped clusters. For $n < 8$, the binding energy of the pure clusters are more than the doped clusters of same size, whereas, for $n > 8$, the trend is reversed. This is because of the absorption of the dangling bonds of the germanium cage by the endohedrally doped TM atom [46]. By adding one germanium atom with $\text{Ge}_{13}(\text{A})\text{TM}$ one can form the optimized ground state structure $\text{Ge}_{14}(\text{A})\text{TM}$ with threefold symmetry in Ge_{14}TM series. Out of the other three optimized geometries, $\text{Ge}_{14}(\text{A})\text{TM}$ is presented in Fig. 1.

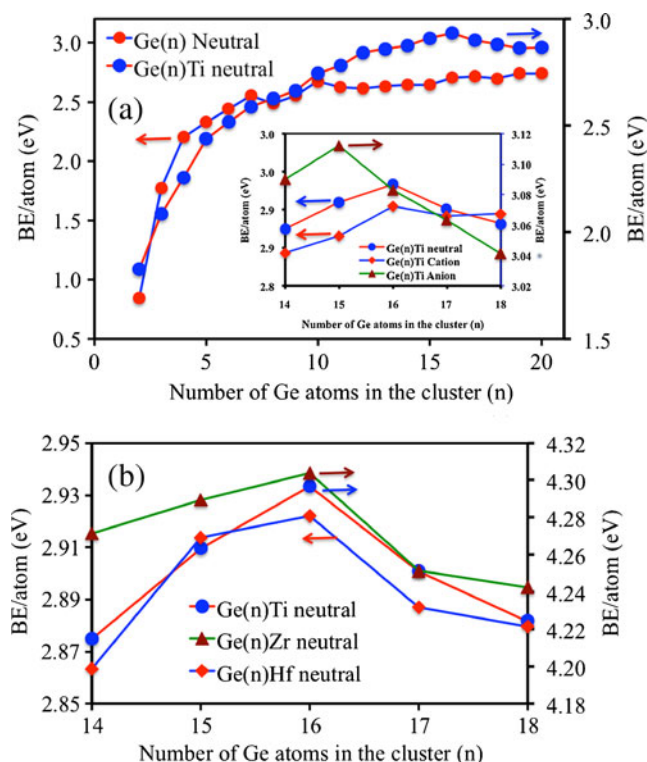


Fig. 2 Variation of binding energy of Ge_nTM with the size of the cluster (n)

The ground state structure $\text{Ge}_{14}(\text{A})\text{TM}$ has threefold symmetry. It contains six pentagons and three isolated rhombi. Again this structure confirms that the structures with pentagons and isolated rhombi are more favorable and follow the isolated rhombus rule [46] in carbon fullerene. The second optimized isomer $\text{Ge}_{14}(\text{B})\text{TM}$ is a symmetrical hexagonal bi-capped structure with total eight rhombi. The optimized ground state structure of $\text{Ge}_{15}(\text{A})\text{TM}$ is quite similar to $\text{Ge}_{14}(\text{A})\text{TM}$. It is easy to understand the $\text{Ge}_{15}(\text{A})\text{TM}$ initial guess structure by adding one germanium atom to the bottom of the rhombi in $\text{Ge}_{14}(\text{B})\text{Ti}$ and then replace Ti by the transition metal atom. The optimized $\text{Ge}_{15}(\text{A})\text{TM}$ structure consists of two connected pentagons and ten rhombi. At this size a total of 11 guess structures are optimized. Out of these 11 structures, only the cage like structures, those which are close to the energy of the ground state structures are presented in Fig. 1. In the next size Ge_{16}TM , ten initial guess geometries are optimized and five are presented. With the increase of the size of the clusters, possibility of searching the ground state structure is always a difficult job. Therefore the numbers of initial guess geometry also increase so that the ground state structures can be obtained. Out of the ten optimized geometries, cage kind five structures are presented in Fig. 1. The geometry of the ground state $\text{Ge}_{16}(\text{B})\text{TM}$ structure can be understand on the basis of the ground state structure of Ge_{12}TM by capping the hexagonal planes using one Ge and a Ge_3 planer triangle

structure. When the hexagonal planes in Ge_{12}Ti structure is capped by two crossed Ge-Ge dimers, optimized $\text{Ge}_{16}(\text{A})\text{Ti}$ structure is obtained. The other structure $\text{Ge}_{16}(\text{C})\text{Ti}$ has two widely separated squares and eight pentagons. Each square is connected to four pentagons separately. This structure is not as symmetrical as found in titanium doped silicon cluster of same size [15, 16]. It is to be noted that with the increase of the size of the clusters beyond $n=16$, distortion starts in the structures and it continues up to the biggest structure for $n=20$ in most of the geometries. In this contest it is worth mentioning that the pure and transition metal doped all Ge_nH_n ($n=6-28$) clusters are very symmetrical [25] because of the strong bonding between the hydrogen and the germanium atoms in the cluster surface. The hydrogen atoms absorb the unsaturated dangling bonds in the cage and give a symmetrical structure. Other bigger clusters in the present study also show distorted structures. Optimized Zr and Hf doped ground state germanium clusters for $n>13$ are presented in Fig. 1. These structures are similar to the Ti doped germanium clusters of same size. In summary, the growth behavior of TM@Ge_n ($\text{TM} = \text{Ti}, \text{Zr}$ and Hf) doped germanium clusters follow two different trends. In lower range of the cluster size ($n<10$), the geometry of the cluster changes from planer to three-dimensional ground state structure where the TM atom absorbs either exohedrally or stays as surface capped element with the germanium clusters. From $n=1-7$, TM elements attached with the germanium clusters as an exohedrally connected element. For $n=8$, the TM atom absorbs partially in optimized ground state cluster $\text{Ge}_8(\text{A})\text{TM}$. In the next stage of growth of the clusters, TM atoms absorb endohedrally in GS state geometry for $n=9$ and then form a cage like cluster for $n>9$ as ground state geometry. The cages are very much symmetric up to $n=16$ and for bigger clusters, the TM atoms stay inside the cage cluster, but the cage gets distorted in some isomers. Depending on the cage size and combination of pentagons, rhombi and hexagons, sometimes in bigger sizes (i.e., $n=24$), it is also possible to get a symmetrical optimized TM doped cage like germanium clusters as ground state [2].

In the next section different physical and chemical properties of the neutral as well as the charged clusters will be discussed.

Electronic structure and stabilities of TM@Ge_n nanoclusters

In the present section the electronic structure and stabilities of the TM@Ge_n ($\text{TM} = \text{Ti}, \text{Zr}$ and Hf) clusters of different size will be discussed by calculating their binding energy (BE), HOMO-LUMO gap (or ΔE), embedding energy (EE), relative stability (Δ_2), ionization potential (IP) etc. Following the behavior of these parameters with the cluster size, we will investigate whether the electron-counting rule can explain the relative stability of the clusters or not.

To explore the relative stability of Ge_nTM clusters with the increasing number of ‘n’ from $n=1$ to 20, binding energy (BE), embedding energy (EE), HOMO-LUMO gap (ΔE), and the relative stability or second order energy difference (Δ_2) of the clusters are studied.

In the present calculation, binding energy per atom following WW rule [30] of the clusters is defined as follows:

$$BE = -(E_{Ge_nTM} - nE_{Ge} - E_{TM}) / (n + 1), \tag{1}$$

where BE is the binding energy per atom in the cluster $TM@Ge_n$, E_{Ge_nTM} , E_{Ge} and E_{TM} are the ground state energies of $TM@Ge_n$ cluster, germanium and transition metal atoms respectively. Following the same expression, binding energy of the anion clusters is also calculated. The variation of binding energy of different clusters both in neutral and charged states are shown in Fig. 2a–b. In Fig. 2a the graphs show the rapid increase in binding energy per atom of the clusters in small size range from $n=1$ to 9 indicating thermodynamic instability of the small size clusters. For the sizes $n>9$ the binding energy curve increase with relatively lower rate. It is to be noted that for $n=16$ and $n=15$ in neutral and anion clusters, binding energy show maximum values in the whole range of study. Both of these clusters are 20-electron cluster. Signature of 20-electron cluster in cationic state is not very clear in the binding energy variation. The relative change in BE per atom in cationic cluster for $n=15-17$ is in the range of ± 0.01 eV or even less. Therefore it is not presented in the variation of other parameters with the size of the cluster. There is no significant change in binding energy in the graphs for $n=14$ (or 13 in anion) and $n=18$ (or 17 in anion) sizes so that these clusters cannot be justified as 18 or 22 electron clusters. Presence of relative peak in the binding energy graphs for $n=12$ could be due to the higher relative stability of the geometry of the cluster. Variation of BE of the other transition metal clusters is shown in Fig. 2b. The nature show that Zr and Hf doped clusters are also show maximum values in binding energy for $n=16$. The next parameter that can indicate the thermodynamic stability of the cluster is embedding energy (EE). In the present study embedding energy of the cluster is defined as:

$$EE = -(E_{Ge_n} + E_{TM} - E_{Ge_nTM}) \tag{2}$$

which, is always negative. As before, E_{Ge_nTM} and E_{Ge_n} are the ground state energy of Ge_nTM and Ge_n clusters, E_{TM} is the energy of the TM atom. The above equation can be modify by imposing WW spin conservation rule [30] as follows:

$$EE^{WW} = {}^M E_{Ge_nTM} - {}^M E_{Ge_n} - {}^0 E_{TM} \tag{3}$$

where M is the total spin of the cluster or of the atom in unit of $\hbar/2$. For example if the cluster is in triplet state, in

calculating the embedding energy we have to take either the energy of the TM atom in triplet state with pure germanium cluster in singlet spin state or the pure germanium cluster in triplet spin state with singlet titanium atom. More precisely, in such case one has to take the value where EE is relatively less. Therefore, in our present calculation, for smaller size clusters where the spin states are not singlet, the EE is calculated according to the above method. It is necessary to mention that, taking the energy of TM atom in triplet state in a series of EE calculation does not make any relative change, because it will shift the graph by an amount of energy $E_{TM}^3 - E_{TM}^1$. So the nature of the graph will remain unchanged. Therefore, whether it is necessary to include WW rule [30] in calculating the relative stability always or to follow different method for different transition metal doped germanium clusters is not possible to say. In the present calculation since there is only one TM atoms from the 2nd column of the periodic table, so WW rule [30] in calculating relative stability is included. For the charged clusters the situation gets a little more complicated. In case of $Ti@Ge_n$ clusters the excess charge can be either on the Ge_n cluster or on the Ti atom. Hence we have to consider both of these possibilities, and take whichever is smaller. In such case EE for anionic can be written as:

$$EE({}^M Ge_n Ti^-) = {}^M E_{Ge_n Ti^-} - {}^1 E_{Ge_n} - {}^{M\pm 1} E_{Ti} \text{ or } {}^M E_{Ge_n Ti^-} - {}^0 E_{Ge_n} - {}^M E_{Ti} \tag{4}$$

Variation of EE with size (n) for both neutral $TM@Ge_n$ and anionic $Ti@Ge_n$ clusters are shown in Fig. 3. In the neutral clusters, EE is minimum at $n=16$. Interestingly, in the anionic clusters, EE has a dip at $n=15$. Thus both in the neutral and cationic series, 20-electron clusters have enhanced stability. The dip in EE graph indicates the most favorable size (here $n=16$) and geometry of pure germanium cluster to incorporate a TM atom. To further check the most stable cluster in the Ge_nTM series during its growth from $n=1$ to $n=20$, Ge atom is added one by one to Ge-TM dimer and the stability parameter is calculated. The stability

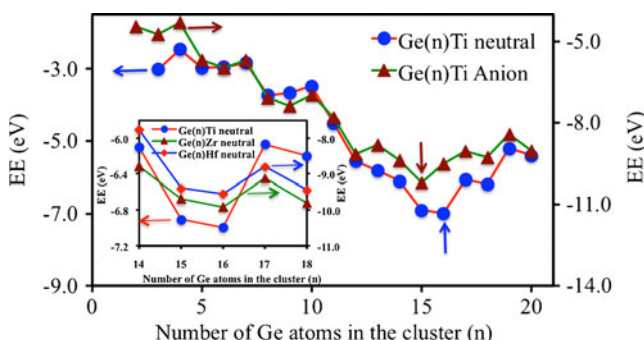


Fig. 3 Variation of embedding energy of Ge_nTM with the size of the cluster (n). In the figure vertical arrows represent the position of 20-electron cluster in the graph

parameter $\Delta_2(n)$ is defined by following previously reported work [11] as follows:

$$\begin{aligned}\Delta_2(n) &= (E_{Ge_{n+1}TM} - E_{Ge_nTM}) - (E_{Ge_nTM} - E_{Ge_{n-1}TM}) \quad (5) \\ &= E_{Ge_{n+1}TM} + E_{Ge_{n-1}TM} - 2E_{Ge_nTM}\end{aligned}$$

According to this definition, large positive values of $\Delta_2(n)$ are indicative of enhanced stability as they correspond to a gain in energy during formation from the preceding size and lower gain in energy to the next size. Variation of stability with the size of the neutral and anionic clusters is shown in Fig. 4. Presence of peak position at $n=16$ and $n=15$ for neutral and anionic clusters respectively are representing the enhanced stability of these clusters. Both of these clusters are again supporting the enhanced stability in 20-electron clusters like binding energy and embedding energy graphs discussed before. In general the clusters which are in magic (positive stability) in neutral states remain magic in most of the negatively charged state.

To see the nature of the growth behavior around $n=16$ cluster which is a 20-electron cluster, fragmentation energy, $\Delta(n, n-1)$ or FE of the clusters in each step starting from Ge-TM dimers is calculated. Here fragmentation energy is defined as follows:

$$\begin{aligned}\Delta(n, n-1) &= -(E_{Ge_nTM} - (E_{Ge_{n-1}TM} + E_{Ge})) \quad (6) \\ &= E_{Ge_{n-1}TM} + E_{Ge} - E_{Ge_nTM}\end{aligned}$$

Figure 5 represents the variation of FE with the cluster size. It is clear that there is a sharp drop in FE from $n=16$ to 17 in neutral state and from $n=15$ to 16 in anionic $Ge_{16}Ti$ cluster. This is an indication of the sudden change in stability from $n=16$ (or $n=15$) to $n=17$ (or $n=16$) in neutral (or anionic) state. Behavior of fragmentation energy variation of Zr and Hf doped germanium clusters within the size range $14 \leq n \leq 18$ is also shown in Fig. 5. In all doped clusters FE drop sharply when it changes from $n=16$

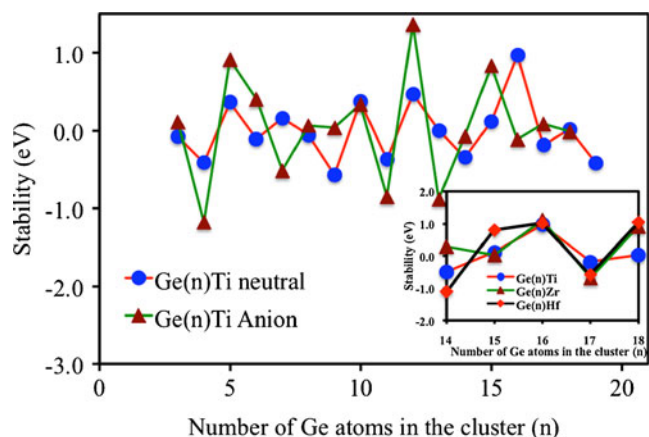


Fig. 4 Variation of stability of Ge_nTM with the size of the cluster (n)

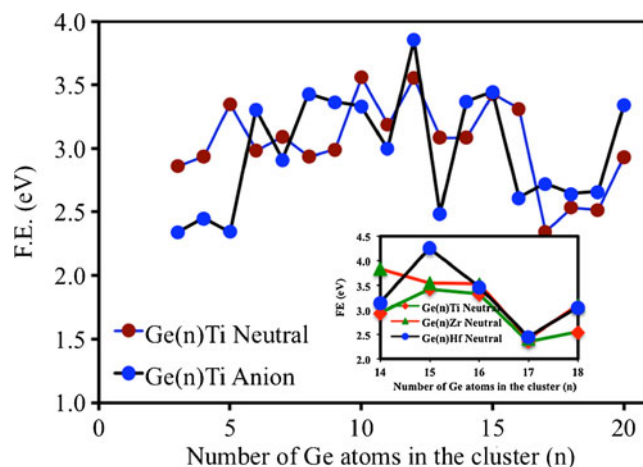


Fig. 5 Variation of fragmentation energy of Ge_nTM with the size of the cluster (n)

to 17. This is an indication of higher stability of the clusters at $n=16$ size. Therefore, it is clear that BE, EE and $\Delta_2(n)$ support the relatively higher thermodynamic stability of the clusters for $n=16$ and 15 in neutral and anionic states respectively.

To get the idea about the kinetic stability of the clusters in chemical reactions the HOMO-LUMO gap (ΔE), ionization potential (IP), electron affinity (EA), chemical potential (μ), chemical hardness (η) and their ratio ($-\mu/\eta$) are calculated. In general with the increase of HOMO-LUMO gap the reactivity of the cluster decrease. HOMO-LUMO gaps of neutral and anionic Ge_nTM clusters are plotted in Fig. 6. Overall, there is a decrease of the HOMO-LUMO gap with the size of the clusters both in neutral and anionic state with some local oscillations over and above the decreasing trend. It is to be noted that these clusters have large HOMO-LUMO gap varies between 1.00 to 3.13 eV. Although there is no sharp global peak as in other parameters as discussed, there is a clear local peak at $n=16$

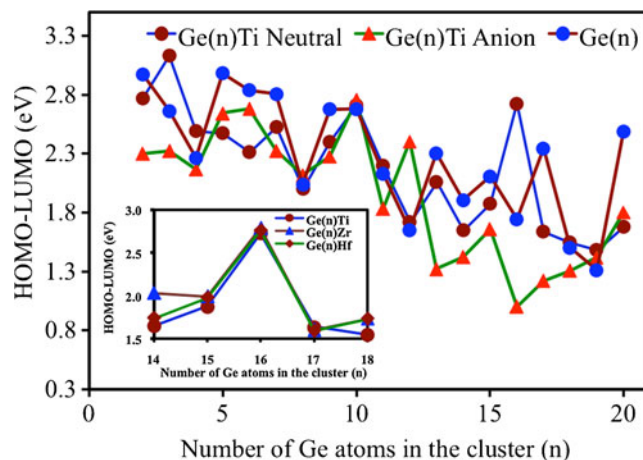


Fig. 6 Variation of HOMO-LUMO gap of Ge_nTM with the size of the cluster (n)

and 15 in neutral and anionic Ge_nTi clusters. This again points to an enhanced stability of 20-electron clusters. It is also worth noticing that at almost all sizes there is a decrease in the HOMO-LUMO gap on TM (Ti, Zr and Hf) encapsulation in pure Ge_n (Fig. 6) clusters in the whole range, but there is a remarkably high value of the gap at $n=16$. In the previous report [15] it is found that the drops are quite marked at $n=5, 6, 9, 14, 15,$ and 17 in Si_nTM ($TM = Ti, Zr$ and Hf) cluster which is different from the present study. Therefore the trend of variation of HOMO-LUMO gap is different in different host for the same transition metal doping.

As mentioned in earlier section, enhanced stability of 20-electron clusters can be rationalized in terms of electronic shell models developed for metal clusters. For metal clusters whenever a new shell starts becoming occupied for the first time, the adiabatic ionization potential (IP) drops sharply [28]. For example, in the report, de Heer [47] has shown that $n=20$ is a shell field configuration for Li_n clusters and there is a sharp drop in ionization potential when the cluster change from Li_{20} to Li_{21} cluster by absorbing a Li atom in it. If the enhanced stability of the 20-electron $Ge_{16}TM$ cluster is due to a shell field configuration then there should be a sharp drop in IP as one more germanium atom is added to it. This is in particular what we see in the IP variation of Ge_nTi cluster when plotted against n as shown in Fig. 7a. There is a peak in IP at $n=16$ and then there is a sharp drop from $n=16$ to 17 and further upto $n=20$ (not shown). Sharp drop in IP from $n=16$ to $n=17$ is perhaps the strongest indication that assumption of a nearly free-electron gas inside the Ge cage is a good model for Ge_nTi clusters, similar to Si_nTM clusters [28]. Variation of IP of the other transition metal clusters is shown in Fig. 7b.

It is well known that chemical stability of the clusters also can be indicated by their electron affinity, which is defined as,

$$EA(eV) = E_{Ge_nTi^-} - E_{Ge_nTi} \quad (7)$$

and is a negative quantity. The clusters those are chemically stable in nature should show relatively low electron affinity. Variation of electron affinity of the Ge_nTi clusters is shown in Fig. 7. It clearly indicating that the electron affinity of the cluster is decreasing for n from $n=14$ and show a minimum value at $n=16$ which is a 20 electron cluster. Then there is a sharp increase (in negative sense) of electron affinity from $n=16$ to 20. Relative dip in EA at $n=16$ is an indication of enhanced stability of $Ge_{16}Ti$ clusters. Variation of EA of the other transition metal clusters is shown in Fig. 7c.

Further, to verify the chemical stability of Ge_nTi clusters, chemical potential (μ) and chemical hardness (η) of the ground state isomers are calculated. In practice chemical potential and chemical hardness can be expressed in terms

of electron affinity and ionization potential. In terms of total energy consideration if $E(N)$ is the energy of the N electron system, then energy of the system containing $N+\Delta N$ electrons where $\Delta N < < N$ can be expressed as:

$$E(N + \Delta N) = E(N) + \left. \frac{dE}{dx} \right|_{x=N} \Delta N + \frac{1}{2} \left. \frac{d^2E}{dx^2} \right|_{x=N} (\Delta N)^2 + \text{Neglected higher order terms} \quad (8)$$

Then, μ and η can be defined as:

$$\mu = \left. \frac{dE}{dx} \right|_{x=N} \quad \text{and} \quad \eta = \left. \frac{dE}{dx} \right|_{x=N} \quad (9)$$

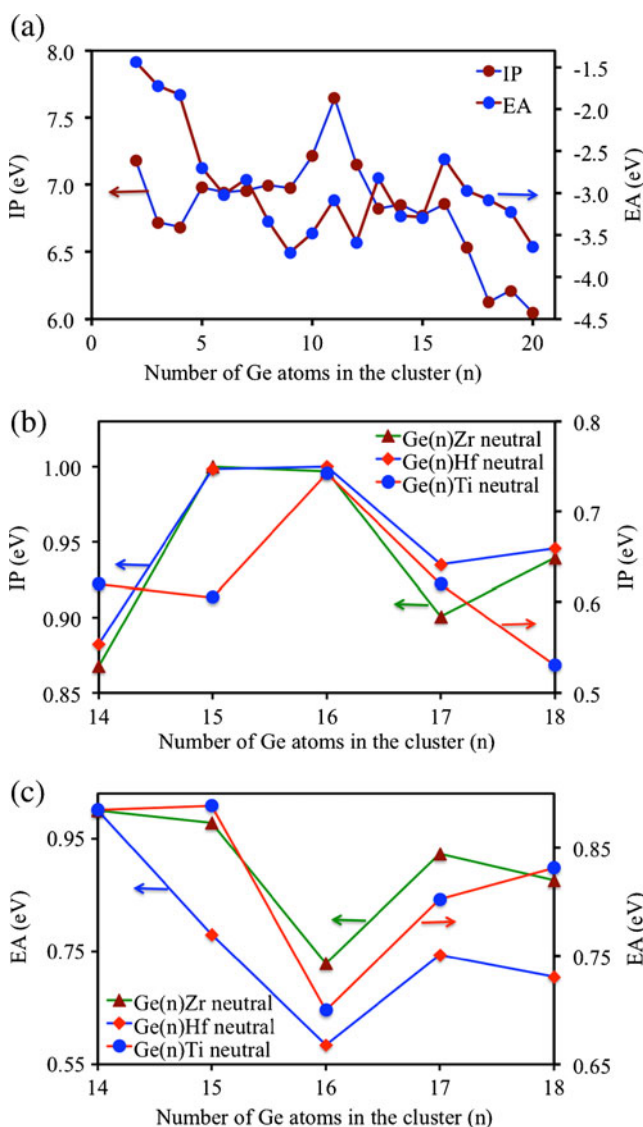


Fig. 7 Variation of IP and EA of Ge_nTi with the size of the cluster (n). All data presented in (b) and (c) are the normalized values of IP and EA in positive scale respectively

Since, $IP = E(N - 1) - E(N)$ and $EA = E(N) - E(N + 1)$. By setting $\Delta N = 1$, μ and η are related to IP and EA via the following relations:

$$\mu = -\frac{IP + EA}{2} \text{ and } \eta = \frac{IP - EA}{2} \quad (10)$$

Consider two systems with μ_i and η_i ($i=1,2$) contracting each other, where some amount of electronic charge (ΔQ) transfer from one to other. The quantity ΔQ and the resultant energy change (ΔE) due to the charge transfer can be determined in the following way:

If $E(N + \Delta Q)$ is the energy of the system after charge transfer then it can be expressed for the two different systems 1 and 2 in the following way:

$$E_1(N_1 + \Delta Q) = E_1(N_1) + \mu_1(\Delta Q) + \eta_1(\Delta Q)^2 \quad (11)$$

and

$$E_2(N_2 - \Delta Q) = E_2(N_2) - \mu_2(\Delta Q) + \eta_2(\Delta Q)^2 \quad (12)$$

Corresponding chemical potential becomes, $\mu'_1 = \left. \frac{dE_1(x+\Delta Q)}{dx} \right|_{x=N_1} = \mu_1 + 2\eta_1 \Delta Q$ and $\mu'_2 = \left. \frac{dE_2(x-\Delta Q)}{dx} \right|_{x=N_2} = \mu_2 - 2\eta_2 \Delta Q$ to first order in ΔM after the charge transfer.

In chemical equilibrium, $\mu'_1 = \mu'_2$ which gives the following expressions:

$$\Delta Q = \frac{\mu_2 - \mu_1}{2(\eta_1 + \eta_2)} \quad \text{and} \quad \Delta E = \frac{(\mu_2 - \mu_1)^2}{2(\eta_1 + \eta_2)} \quad (13)$$

In the expression, ΔE is the energy gain by the total system (1 and 2) due to exclusive alignment of chemical potential of the two systems at the same value. From the above expression of ΔQ and ΔE it is clear that for easier charge transfer from one system to another it is necessary to have a large difference in μ together with low η_1 and η_2 . Therefore, ΔQ and ΔE can be taken as the measure factors to get the idea about the reaction affinity between two systems. Since they are function of the chemical potential and chemical hardness related to the system, so it is important to calculate these parameters of a system to know about its chemical stabilities in a particular environment.

Keeping the above preliminaries in mind, chemical potential (μ) and chemical hardness (η) for Ti doped Ge_n clusters is calculated. To search for the chemical stability of a cluster in chemical reaction or ability to accept electron with low chemical hardness, the $|\mu/\eta|$ ratio along with the chemical potential and chemical hardness is calculated (both of these parameters is normalized to 1) and presented in Fig. 8 with the variation of the size of the cluster. In the Fig. 8 these three parameters are plotted in positive sense. Peak at $n=16$ in normalized chemical potential plot is

actually indicating low value of chemical potential and hence low affinity of the system to take part in chemical reaction in a particular environment. Again at $n=16$, the presence of a localized dip in normalized chemical hardness plot is also supporting the result of low chemical affection Ti@Ge_{16} cluster. The ratio of these two parameters in positive sense shows a peak and hence indicating the low chemical affinity. Since $n=16$ is a 20-electron cluster, it is clear that this cluster also show very low affinity in chemical reaction and is in stability agreement with the other parameters.

To get the idea about the vibrational and optical properties of the TM@Ge_n clusters, vibrational (IR) and Raman spectra of optimized ground state isomers are calculated. Some selected optimized ground state clusters for $n=15, 16$, and 17 are presented in Fig. 9. In the figure the cluster geometry and corresponding most dominating frequencies of IR and Raman spectra are also presented. The absence of any imaginary frequency in the spectrum represents the real nature of the clusters. Careful observation shows the presence of two types of frequencies in IR and Raman spectra of the clusters. The dominant peaks shown in the IR spectra of different clusters are due to the vibration of the germanium atoms in the cage with the transition metal atom as a coupled oscillation, whereas dominant peak in Raman spectrum is only due to the vibration of germanium atoms in the cage. This mode is known as breathing mode in the Raman spectrum where all germanium atoms vibrate in phase. Comparing this with the binding energy graph, it can be seen that the increase in average bond energies (not presented) helps to increase the binding energy per atom in the clusters. It is observed that for both IR and Raman spectra for $n=15$ and 16 are very clear and a number of modes present also much smaller than TM@Ge_{17} clusters. This is because of the symmetrical nature of $n=15$ and 16 clusters. It is observed that in a particular size the dominating frequency of IR spectrum decreases from Ti to Hf. It is also observed that the frequency of the

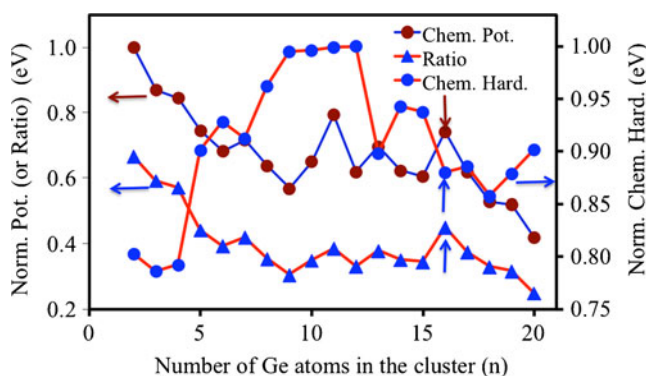


Fig. 8 Variation of chemical potential, chemical hardness and their ratio of Ge_nTi with the size of the cluster (n). In the figure vertical arrows represent the position of 20-electron clusters

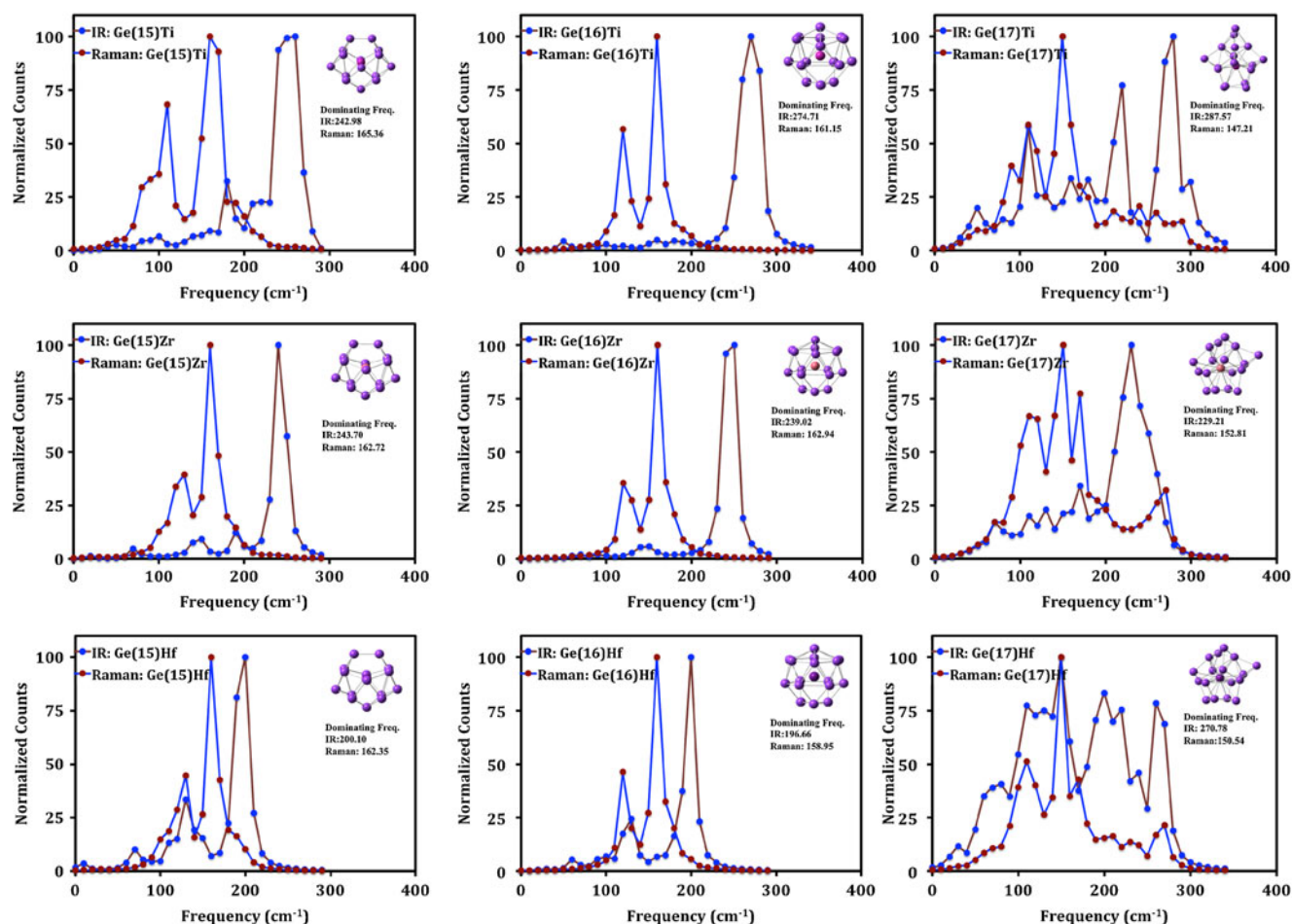


Fig. 9 IR and Raman spectrum of some selected TM@Ge_n clusters

dominating mode in IR spectrum is always much higher than the dominating mode present in Raman spectrum of a particular cluster within the size range $n=15$ to 17 . In a particular size maximum variation of the frequency of the dominating mode in Raman spectrum is 147 cm^{-1} to 166 cm^{-1} , but this range is comparatively much higher in IR spectrum.

Conclusions

In summary, the study of electronic structures and stabilities of transition metal doped germanium nanoclusters is presented. Different properties of neutral and anionic TM-doped optimized geometries of the clusters, like binding energies, stabilities, HOMO-LUMO gap, electron affinities, ionization potentials, charge transfer, infrared and Raman spectra are discussed. Based on the results, the following conclusions can be made:

a) It is always favorable to attach a TM-atom to Ge clusters at all sizes, as the EE turns out to be negative

(as according to the definition of the EE in the present study) in every case of neutral and anionic clusters. All clusters with size $n>9$ absorb TM atom endohedrally in the cage of Ge_n pure cluster. In all TM-doped clusters beyond $n>7$, the spin on the TM atom is quenched.

b) More interesting is the relative stability of these clusters. As measured by their BE, EE, $\Delta(n,n-1)$ and $\Delta_2(n)$, both neutral and anionic clusters having 20 valence electrons show enhanced stability, in agreement with shell model predictions. It also shows up in the IP values of the Ge_nTi clusters, as there is a sharp drop in IP from $n=16$ to 17 . Validity of nearly free-electron shell model is similar to that of transition metal doped silicon clusters. While Ge₁₆Ti is particularly a stable species, Ge₁₅Ti with its smaller IP may form ionic compounds with halogen atoms. Although the signature of stability is not so sharp in the HOMO-LUMO gaps of the anionic cluster, but there is still a local maximum at $n=15$, indicating enhanced stability of a 20-electron cluster, but this signature is very prominent for $n=16$ in neutral cluster which is also a 20-electron cluster. Other parameters like, EA, chemical

potential are related to the chemical stabilities of the neutral cluster for $n=16$ also supports the identical stability nature of the clusters. Identification of such stable species, and variation of chemical properties with size in the transition metal doped germanium clusters will help to design germanium-based superatoms. The present work is the preliminary step in the direction of designing of transition metal doped germanium superatoms and it will be followed by more detailed studies on different other systems with bigger cluster sizes.

- c) Infrared intensities and Raman activities show distinct spectra for different optimized clusters. This also reflects the change of bonding nature with the size of the clusters. This can be useful to identify the structures of these clusters from experiments. IR and Raman spectra do not show any imaginary frequency in optimized ground state. Moreover, it shows that the number of modes present in the IR and Raman spectrum in symmetric clusters is less compared to the less symmetric clusters. Since there are different dominating frequencies present in IR range of the vibrational spectrum of the cluster which is a function of the size and the nature of transition metal doping, therefore this data is very much important to design a transition metal doped hybrid germanium IR sensitive nano device or helpful to design other types of germanium-transition metal based nano devices.

Acknowledgments Complete computations using Gaussian 03 were performed at the cluster computing facility, Harish-Chandra Research Institute, Allahabad, UP, India (<http://cluster.hri.res.in>).

References

- Ho KM, Shvartsberg AA, Pan B, Lu ZY, Wang CZ, Wacker JG, Fye JL, Jarrod MF (1998) *Nature* 392:582–585
- Wang J, Chen X, Liu JH (2008) *J Phys Chem A* 112:8868–8876
- Zhao WJ, Wang YX (2008) *Chem Phys* 352:291–296
- Jarrold MF, Constant VA (1991) *Phys Rev Lett* 67:2994–2997
- Benedict LX, Puzer A, Williamson AJ, Grossman JC, Galli G, Klepeis JE, Raty JY, Pankratov O (2003) *Phys Rev B* 68:85310–85317
- Brown WL, Freeman RR, Raghavachari K, Schluter M (1987) *Science* 235:860–865
- Hiura H, Miyazaki T, Kanayama T (2001) *Phys Rev Lett* 86:1733–1736
- Hayashi S, Kanzaya Y, Kataoka M, Nagarede T, Yamamoto K (1993) *Z Phys D Atom Mol Cl* 26:144–146
- Bandyopadhyay D, Kaur P, Sen P (2010) *J Phys Chem A* 114:12986–12991
- Polman A (2002) *Nat Matters* 1:10–12
- Bandyopadhyay D, Sen P (2010) *J Phys Chem A* 114:1835–1842
- Jarrold MF, Bower JE (1992) *J Chem Phys* 96:9180–9190
- Kumar V, Kawazoe Y (2001) *Phys Rev Lett* 87:045503–045506
- Kumar V, Kawazoe Y (2002) *Phys Rev Lett* 88:235504–235507
- Bandyopadhyay D (2008) *J Appl Phys* 104:084308–084314
- Bandyopadhyay D (2009) *Mol Simul* 35:381–394
- Kumar M, Bandyopadhyay D (2008) *Chem Phys* 353:170–176
- Beck SM (1987) *J Chem Phys* 87:4233–4234
- Beck SM (1989) *J Chem Phys* 90:6306–6312
- Ohara M, Miyajima K, Pramann A, Nakajima A, Kaya K (2002) *J Phys Chem A* 106:3702–3705
- Han JG (2000) *Chem Phys Lett* 324:143–148
- Wang JL, Wang GH, Zhao JJ (2001) *Phys Rev B* 64:205411–305415
- Hou XJ, Gopakumar G, Lievens P, Nguyen MT (2007) *J Phys Chem A* 111:13544–13553
- Negishi Y, Kawamata H, Hayase T, Gomei T, Kishi R, Hayakawa F, Nakajima A, Kaya K (1997) *Chem Phys Lett* 269:199–207
- Bandyopadhyay D (2009) *Nanotechnology* 20:275202–275213
- Huheey JE, Keiter EA, Keiter RL (2000) *Inorganic Chemistry: principles of structure and reactivity*, 4th edn. Harper-Collins College Publisher, New York
- Sen P, Mitas L (2003) *Phys Rev B* 68:155404–155407
- Reveles JU, Khanna SN (2005) *Phys Rev B* 72:165413–165418
- Guo LJ, Zhao G, Gu Y, Liu X, Zeng Z (2008) *Phys Rev B* 77:195417–195424
- Wigner E, Witmer EE (1928) *Z Phys* 51:859–886
- Koyasu K, Akutsu M, Mitsui M, Nakajima A (2005) *J Am Chem Soc* 127:4998–4999
- Kumar V (2003) *Eur Phys J D* 24:227–232
- Burke K, Perdew JP et al. (1998) In: Dobson JF, Vignale G, Das MP (eds) *Electronic Density Functional Theory: Recent Progress and New Directions*. Plenum
- Perdew JP (1991) In: Ziesche P, Eschrig H (eds) *Electronic Structure of solids '91*. Akademie, Berlin
- Becke AD (1988) *Phys Rev A* 38:3098–3100
- Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37:785–789
- Wang J, Han GJ (2005) *J Chem Phys* 123:064306–064321
- Han JG, Hagelberg F (2001) *J Mol Struct THEOCHEM* 549:165–180
- Guo P, Ren ZY, Wang F, Bian J, Han JG, Wang GH (2004) *J Chem Phys* 121:12265–12275
- Guo LJ, Liu X, Zhao GF, Luo YH (2007) *J Chem Phys* 126:234704–234710
- Nagendran S, Sen SS, Roesky HW, Koley D, Grubmüller H, Pal A, Herbst-Irmer R (2008) *Organometallics* 27:5459–5463
- Khon W, Sham LJ (1965) *Phys Rev* 140:A1133–A1138
- Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu B, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (2004) *Gaussian 03, Revision E.01*. Gaussian, Inc, Wallingford, CT
- Lu J, Nagase S (2003) *Chem Phys Lett* 372:394–398
- Kumar V, Kawazoe Y (2002) *Appl Phys Lett* 80:859–861
- Kumar V, Kawazoe Y (2007) *Phys Rev B* 75:155425–155435
- de Heer WA (1993) *Rev Mod Phys* 65:611–676