ORIGINAL PAPER

Tetrahydroisoquinolines acting as dopaminergic ligands. A molecular modeling study using MD simulations and QM calculations

Sebastián Andujar · Fernando Suvire · Inmaculada Berenguer · Nuria Cabedo · Paloma Marín · Laura Moreno · María Dolores Ivorra · Diego Cortes · Ricardo D. Enriz

Abstract A molecular modeling study on 16 1-benzyl tetrahydroisoquinolines (BTHIQs) acting as dopaminergic ligands was carried out. By combining molecular dynamics simulations with ab initio and density functional theory (DFT) calculations, a simple and generally applicable procedure to evaluate the binding energies of BTHIQs interacting with the human dopamine D2 receptor (D2 DR) is reported here, providing a clear picture of the binding interactions of BTHIQs from both structural and energetic viewpoints. Molecular aspects of the binding interactions between BTHIQs and the D2 DR are discussed in detail. A significant correlation between binding energies obtained from DFT calculations and experimental pK_i values was obtained, predicting the potential dopaminergic effect of non-synthesized BTHIQs.

Keywords 1-Benzyl-THIQ · Halogenated-1-benzyl-THIQ · D2-dopamine receptor · Structure-activity relationship · Molecular dynamic simulation · Ab initio and DFT calculation

S. Andujar · F. Suvire · R. D. Enriz (✉)
Departamento de Química, Universidad Nacional de San Luis,
San Luis, Argentina
e-mail: denriz@unsl.edu.ar

S. Andujar · R. D. Enriz
IMIBIO-SL (CONICET),
Chacabuco 915,
5700, San Luis, Argentina

I. Berenguer · N. Cabedo · P. Marín · L. Moreno ·
M. Dolores Ivorra · D. Cortes
Departamento de Farmacología, Facultad de Farmacia,
Universidad de Valencia,
46100 Burjassot Valencia, Spain

Introduction

The dopamine D2 receptor (D2 DR) has been implicated in the mechanism of drugs used in the treatment of disorders such as schizophrenia and Parkinson's disease. For these reasons, a great deal of research has focused on the discovery of novel dopaminergic ligands as potential drug candidates [1]. DR can be classified into two pharmacological families (D1 and D2-like) that are encoded by at least five genes. Which receptor(s) needs to be activated to obtain therapeutic effects in Parkinson's disease has been the subject of controversy [2]. The D2-like DR show high affinities for drugs (antagonists) used in the treatment of schizophrenia (antipsychotics) and those (agonists) utilized to treat the Parkinson's disease [3].
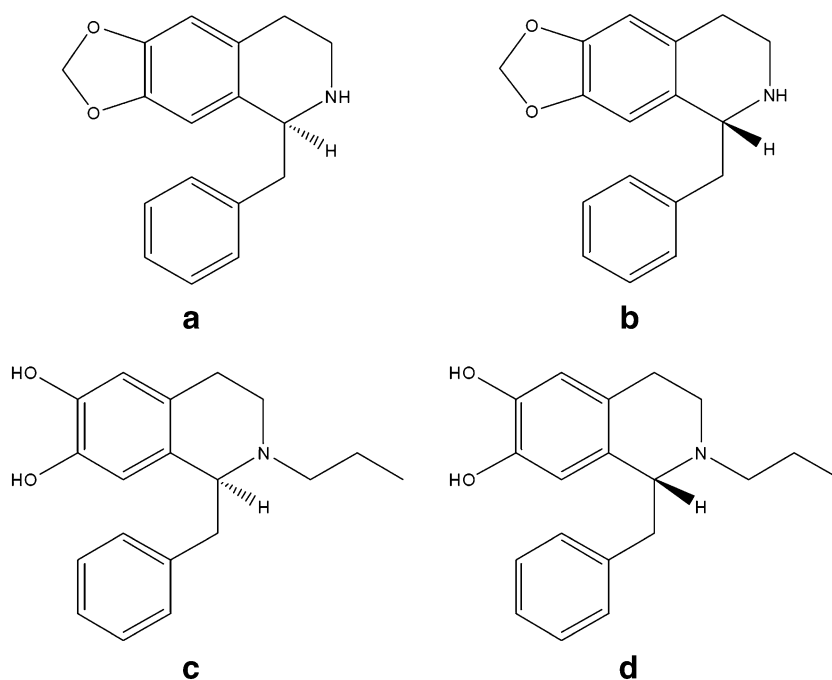
Tetrahydroisoquinolines (THIQs)—the most numerous naturally occurring alkaloids—include 1-benzyl-THIQs and aporphines, both of which have structural similarities to dopamine and can interact with DR [4]. Previous results in our group suggested that some natural and synthetic 1-benzyl-THIQs alkaloids were able to bind to DR [5–7]. In this way, we described the enantioselective syntheses of pairs of dopaminergic (1S)- and (1R)-benzyl-THIQs using (R)- and (S)-phenylglycinol as the chiral source, and we observed that, in these series of 1-benzyl-THIQs, (1S)-enantiomers were 5–15 times more effective at D1-like and D2-like dopamine receptors than (1R)-enantiomers [8] (Fig. 1). On the other hand, we described the preparation in a 'one-pot' sequence of 1-cyclohexylmethyl 7,8-dioxygenated-THIQ, substituted and unsubstituted in the C ring by application of the photo–Fries transposition, followed by a tandem reduction-cyclization and further reduction. Indeed, we accomplished for the first time a regioselective hydrogenation of the benzyl ring in the THIQ system. All 1-cyclohexylmethyl THIQs studied in
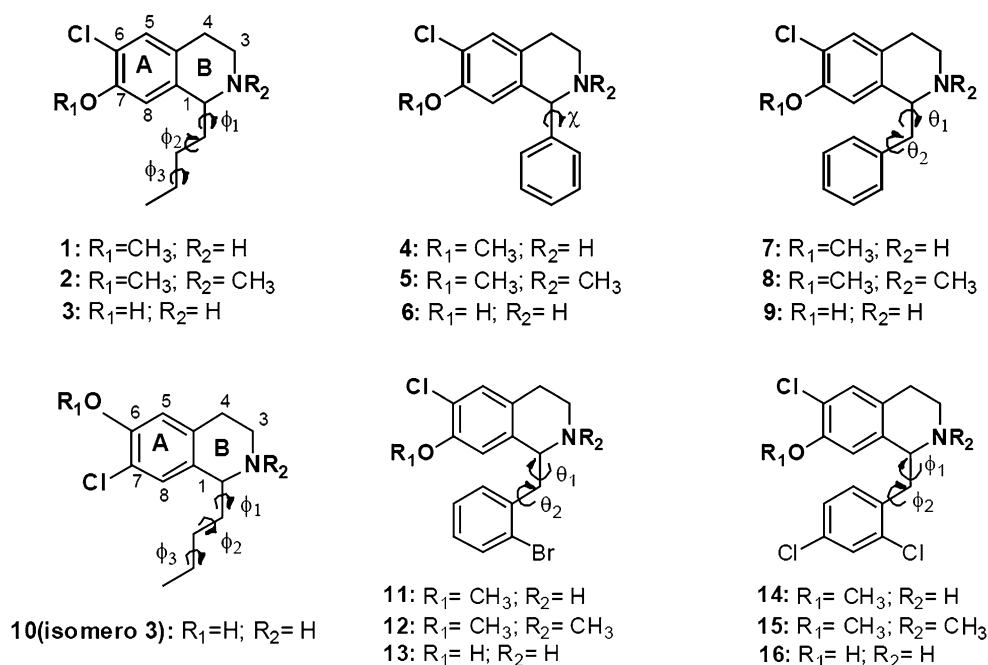
this work were able to displace the D2-like DR radioligand from its specific binding sites in rat striatal membranes, while the *N*-methylated derivatives also showed affinity for the D1-like DR. Recently, we reported the influence of the substitution at the 1-position over a 7-chloro-6-hydroxy-THIQ core [9, 10]. In previous works [8, 11–15], we determined the role of certain structural requirements for improving the affinity for D1 and D2-like DR. Thus, we are able to postulate that the presence of a hydroxyl (OH) and a halogen group (Cl) in the THIQ A-ring could lead to obtaining molecules that can bind selectively to one of the two groups of receptors mentioned above [13, 14]. Preserving the chlorine and hydroxyl (or methoxyl) groups at the C-6 and C-7 positions of the THIQ A-ring, respectively, with a secondary (*N*H) or a tertiary (*N*Me) amine, we explored the impact of inclusion of aliphatic and aromatic groups such as butyl-, phenyl-, benzyl-[8] as well as halogenated-1-benzylbenzyl moieties at 1-position [9] to determine their influence over dopaminergic activity. Thus, we have recently reported five series of 1-substituted-THIQs: 1-butyl-THIQs (compounds **1–3** in Fig. 2), 1-phenyl-THIQs (**4–6**), 1-benzyl-THIQs (**7–9**), 2′-bromobenzyl-THIQs (**11–13**) and 2′,4′-dichlorobenzyl-THIQs (**14–16**). Compound **10**, which was obtained from a fortuitous synthesis, was also included in such a report [8]. During a Bischler-Napieralski cyclization, we observed the same fact reported by Doi et al. in 1997 [16] when preparing 1-butyl-THIQs. The need to add P_2O_5 (and $POCl_3$ at a molar ratio of 1:1) to the cyclodehydration reaction, because of the difficulty of cyclizing the amide when there is a chlorine in the structure (originally at the C-6 position of the A-ring), causes an aberrant cyclization, by

means of formation of a nitrilium intermediate, which gives two positional isomers, clearly identified after the reduction step: 6-chloro-7-hydroxy-1-butyl-THIQ (compound **2**), and 6-hydroxy,7-chloro-1-butyl-THIQ (compound **10**: unexpected cyclization product), in a 1:2 ratio.

All these compounds were assayed in vitro for their ability to displace selective radioligands of D1 and D2 DR from their respective specific binding sites in rat striatal membranes, and were tested for their ability to inhibit in vitro ^{3}[H]-dopamine uptake in rat striatal synaptosomes. Many of these compounds were able to displace both ^{3}[H]-SCH 23390 and ^{3}[H]-raclopride at nano or micromolar (nM or μM) concentration from their specific binding sites in rat striatum, but all compounds had only low or no effect on ^{3}[H]-dopamine uptake [8, 9]. The replacement at the C-1 position of THIQs, is an important factor modulating the selectivity at DR. Compounds **1**, **3**, **10** and **11** (Table 1) show a greater affinity towards D2 receptors when a butyl or a benzyl moiety, respectively, is located in that position. The different activities and selectivity obtained for these compounds can be explained by the different spatial orientations adopted by the varied hydrophobic portions located at C-1, which could give different molecular interactions with the D1 and D2 receptors. Since some BTHIQs have shown a great affinity for the D2 DR, considerable interest has developed in delineating the portions of the BTHIQ molecular structure responsible for its dopaminergic properties and interactions with the D2 DR. The process of drug design could be considerably improved if receptors and their mode of interaction with ligands were known in precise molecular detail. Such

Fig. 2 Structural features of the 16 BTHIQs reported here, showing the different torsional angles

1: R$_1$=CH$_3$; R$_2$= H
2: R$_1$=CH$_3$; R$_2$= CH$_3$
3: R$_1$=H; R$_2$= H

4: R$_1$= CH$_3$; R$_2$= H
5: R$_1$= CH$_3$; R$_2$= CH$_3$
6: R$_1$= H; R$_2$= H

7: R$_1$=CH$_3$; R$_2$= H
8: R$_1$=CH$_3$; R$_2$= CH$_3$
9: R$_1$=H; R$_2$= H

10(isomero 3): R$_1$=H; R$_2$= H

11: R$_1$= CH$_3$; R$_2$= H
12: R$_1$= CH$_3$; R$_2$= CH$_3$
13: R$_1$= H; R$_2$= H

14: R$_1$= CH$_3$; R$_2$= H
15: R$_1$= CH$_3$; R$_2$= CH$_3$
16: R$_1$= H; R$_2$= H

information could then be used to design more defined structures in which the pharmacophoric groups are oriented in the appropriate spatial arrangement for optimal receptor interaction.

In the present work, we report a molecular modeling study performed on 16 BTHIQs acting as dopaminergic ligands. Combined molecular dynamics (MD) simulations and quantum mechanics (semiempirical, ab initio and DFT) calculations were employed in our study to evaluate the molecular

interactions between the BTHIQs and the D2 DR. An excellent correlation between binding energies obtained from DFT calculations and experimental pKi was obtained.

Materials and methods

Theoretical calculations were carried out in two steps. In a first step, we performed MD simulations of the molecular

Table 1 Relative binding energies obtained for the different complexes. Previously reported experimental pKi data are shown in the last column

Compound	Relative binding energy (BE) (kcal/mol)				Specific-D2 ligand [³H]-raclopride
	EU (RHF/6-31G(d))	Δ EU (RHF/6-31G(d))	EU (B3LYP/6-31G(d,p))	Δ EU (B3LYP/6-31G(d,p))	pKi
1	-98.91	22.16	-114.29	24.58	6.108±0.165 [9]
2	-80.00	41.07	-93.62	40.25	5.424±0.026 [9]
3	-111.92	9.15	-135.43	3.44	7.117±0.151 [9]
4	-75.46	45.61	-94.89	43.98	5.212±0.124 [9]
5	-83.02	38.05	-103.00	35.87	5.670±0.406 [9]
6	-90.48	30.59	-105.35	33.52	5.950±0.198 [9]
7	-99.94	21.13	-118.48	20.39	6.014±0.049 [9]
8	-84.29	36.78	-99.93	38.94	5.816±0.181 [9]
9	-116.35	4.72	-137.83	1.04	7.178±0.091 [9]
10	-114.03	7.04	-138.43	0.44	7.220±0.139 [9]
11	-100.47	20.6	-128.07	10.8	6.630±0.092 [10]
12	-81.59	39.48	-113.31	25.56	5.896±0.099 [10]
13	-121.07	0	-138.87	0	7.391±0.139 [10]
14	-74.61	46.46	-101.31	37.56	5.507±0.105 [10]
15	-77.31	43.76	-101.10	37.77	5.230±0.096 [10]
16	-102.79	18.28	-131.14	7.73	6.996±0.105 [10]

interactions between compounds **1**–**16** and D2 DR. In the second step, reduced model systems were optimized using quantum mechanics calculations. Semiempirical (AM1) combined with ab initio [RHF/6-31G(d)] and B3LYP [6-31G(d,p)] calculations were employed in these optimizations.

Molecular dynamics simulations

It must be pointed out that the principal goal of the MD simulations performed here was not to obtain a new D2 DR by homology. Our aim in this study was less ambitious; we wished to obtain a reasonable indication of the relationship between the structures of compounds **5**–**7** and their potential affinities for the binding pocket of D2 DR. Thus, for this purpose, we considered it more appropriate to use a previously reported and extensively tested model for D2 DR [17]. In fact, there are many molecular modeling studies in the literature reporting D2 DRs obtained by homology, all of them structurally closely related [18–20]. Thus, in the present study, we used the D2DR model previously reported in reference [17]. The ligand topologies were built using the mktop program [21]. For this purpose, we used the previously optimized geometry at RHF/6-31G (d) level of theory of the global minimum of each ligand. In the present study, we used an approach where manual docking was guided by information from site-directed mutagenesis and short docking simulations, with both the receptor and the ligand free to move. Structurally similar parts of the ligands were oriented in similar positions in the receptor model, which was described by Mansour et al. [22] and Lan et al. [23]. Thus, receptor–ligand complexes were prepared in order to obtain the input files for MD runs. Several docking positions were considered and the strongest receptor interactions were examined in detail.

The MD simulations and analysis were performed using the GROMACS 3.2.1 simulation package [24, 25] with the OPLS-AA force field [26–30] and the rigid SPC water model [31, 32] in a cubic box with periodic boundary conditions. Receptor–ligand complexes were embedded in a box containing the SPC water model that extended to at least 1 nm between the receptor and the edge of the box, resulting in a box of 7.17 nm in side length. The total number of water molecules was 11,330 for the different simulations. Three Na^+ ions were then added to the systems by replacing water in random positions, thus making the whole system neutral. The time step for simulations was 0.001 ps for a complete simulation time of 5 ns. For long-range interactions, the particle-mesh Ewald (PME) [33–35] method was used with a 1 nm cut-off and a Fourier spacing of 0.12 nm. The MD protocol consisted of several preparatory steps: energy minimization using the conjugate gradient model [36, 37] density stabilization (NPT conditions), and finally production of the MD simulation

trajectory. All production simulations were performed under NVT conditions at 310 K, using Berendsen's coupling algorithm [38] for keeping the temperature constant. The compressibility was 4.8×10^{-5} bar^{-1}. All coordinates are saved every 5 ps. The SETTLE [24] algorithm was used to keep water molecules rigid. The LINCS [39] algorithm was also used to constrain all C-α atom positions for the receptor in order to avoid unfolding problems. The simulations were analyzed using the analysis tools provided in the Gromacs package.

Histidine in the active site is a potential problem because the state of His (neutral or protonated) is a controversial topic. We were particularly interested in performing simulations under physiological conditions (pH≈7). Previous reports have indicated that, under physiological conditions (pH≈7), histidine located in a hydrophobic environment (hydrophobic pocket without water molecules) is in neutral form [40]. In addition, previous simulations performed for D3DR by Micheli et al. [20] also considered the histidine residue to be neutral. Thus, on the basis of these results, we considered His in neutral form in our calculations. This amino acid was calculated as follow: protons were added using the program pdb2gmx, in the GROMACS suite of programs, for optimization of the hydrogen bond network. His protons were placed by default; these selections were done automatically (His was in neutral form). This is based on an optimal hydrogen bonding conformation. Hydrogen bonds are defined based on simple geometric criteria, specified by the maximum hydrogen–donor–acceptor angle and donor–acceptor distance.

It should be noted that the compounds reported here possess one chiral center, and are therefore enantiomeric with the possibility of two isomers (*1-S* and *1-R*). However, we did not perform an enantiomeric resolution for previously reported biological assays; thus, only one isomer of each compound was evaluated in our MD simulations. To choose the isomeric forms of each compound, we considered on the one hand previously reported results [15] and, on the other, preliminary and specially performed exploratory simulations determining the spatially preferred form of each compound (results not shown). Our previous experimental results on structurally related compounds suggested that the *S* form would be the preferred isomer for these compounds [15]. The preliminary and exploratory MD simulations are in agreement with these experimental data, indicating that the spatial ordering adopted by *1-S* forms gives adequate orientation of the molecules to interact in the active site of the dopamine D2 receptor.

The equilibrium state of the complexes was observed from the onset of simulation until 5 ns. The temperature was stabilized at 310±4 K for all complexes. The potential energy stabilized in a short time period (around 0.5 ns), and

the values obtained suggested that the system was well equilibrated.

Considering the 5 ns of MD simulation, and from the time profiles, it was concluded that some properties of the ligand–receptor complexes reached stable average values at around 0.5 ns, whereas others take longer time periods. For this reason, and to ensure full equilibration, only the last 4.5 ns were taken into account for the analysis. After discarding the first 0.5 ns of the trajectory, we followed the changes in spatial ordering of the ligand–receptors complexes.

Quantum mechanics calculations

The binding pocket of the D2 L–R (ligand-receptor) was defined according to Teeter et al. [41] and Neve et al. [42]. In our reduced model system, only 13 amino acids were included in molecular simulations. The size of the molecular system simulated and the complexity of the structures under investigation restricted the choice of the quantum mechanical method to be used. Consequently, the semiempirical AM1 method was selected combined with ab initio calculations (RHF/6-31G(d)). The torsional angles of the ligands and the flexible side-chains of the amino acids as well as the bond angles and bond lengths of the moieties involved in the potential intermolecular interactions were optimized at the semiempirical level. Next, the torsional angles of the ligands and the flexible side-chains of the amino acids as well as the potential intermolecular interactions were optimized at RHF/6-31G(d) and DFT [B3LYP/6-31G(d,p)] levels of theory. In contrast, the torsional angles of backbones as well as the bond angles and bond lengths of non-interacting residues were kept frozen during the calculations.

The binding energy of the complexes was calculated, with the approximation neglecting the superimposition of error due to the difference between the total energies of the complex with the sum of the total energies of the components:

$$BE_{QM} = E_{L/D2DR} - (E_{D2DR} + E_L) \tag{1}$$

where BE_{QM} is the binding energy, $E_{L/D2DR}$ is the complex energy, E_{D2DR} the energy of the reduced receptor model (binding pocket) and E_L the energy of the ligand.

All the quantum mechanical calculations reported here were carried out using the Gaussian 03 program [43].

Spatial views shown in Figs. 3, 9 and 10 were constructed using the UCSF Chimera program [44] as the graphic interface.

Results and discussion

Our molecular modeling study was carried out in two steps. First, we performed MD simulations of the molecular interactions between the compounds shown in Fig. 2 with the human D2 DR (Fig. 3). In the second step, reduced model systems (shown as a circle in Fig. 3) were optimized using quantum mechanic calculations. Semiempirical (AM1) combined with ab initio [RHF/6-31G(d)] and DFT [B3LYP/6-31G(d,p)] calculations were employed for these optimizations.

Molecular dynamic simulations

Comparing the results obtained for the different complexes led to interesting general conclusions. Consistent with previous experimental [22] and theoretical [45] results, our simulations indicate the importance of the negatively charged aspartate 86 for binding of these ligands. A highly conserved aspartic acid (Asp 86) in trans-membrane helix 3 (TM3) is important for the binding of both agonists and antagonists to the D2 receptor, [22, 46, 47], and its terminal carboxyl group may function as an anchoring point for ligands with a protonated amino group [23, 41, 42, 47]. In the present study, all the compounds simulated were docked into the receptor with the protonated amino group near Asp 86. After 5 ns of MD simulations, the ligands had moved slightly but in a different form compared with the initial position. However, the strong interaction with Asp 86 was maintained for all complexes (see Fig. 4), supporting the suggestion that Asp 86 functions as an anchoring point for ligands with a protonated amino group.

Pharmacological data [22, 48] indicate that the hydroxyl groups of dopaminergic ligands are of primary importance in stabilizing binding, suggesting that the serine residues (141 and 144) of the D2 receptor may not be equally important for binding affinity. Individual mutation of serines 141 and 144 in TM5 to alanine produced asymmetrical effects on dopamine receptor binding. These results indicated that Ser 141 might be differentially important for dopamine binding. In addition, site-directed mutagenesis studies have indicated that a cluster of serine residues in TM5 (Ser 141, Ser 144) and in TM4 (Ser 122 and Ser 118) is important for agonist binding and receptor activation [45, 47–49]. It was suggested that the serine cluster and dopamine form a hydrogen-bonding network. Such a hydrogen-bonding network was reproduced by the MD simulation of these complexes (Fig. 5). In these complexes, the strongest contributor to the network was Ser 141, which is consistent with the experimental observation that a Ser 141 Ala mutated receptor completely lost dopamine-induced activation [22]. The 7-hydroxyl group of compound **3** displayed another significant hydrogen bond interaction with Ser 122; however, this interaction is weaker with respect to the hydrogen bond with Ser 141.

Figure 5 shows that compounds **3** and **9** display strong hydrogen bond interactions with Ser141 during the entire

Fig. 3 Spatial view obtained for the dopamine D2 receptor (D2 DR) model. The plot was performed using the UCSF Chimera program [44] program as a graphic interface. Conformations used as starting geometries for the molecular dynamics (MD) simulations of compounds **3** (*cyan*), **6** (*green*) and **9** (*yellow*) are shown. The binding pocket optimized from quantum mechanics calculations is denoted with a *circle*. The numbers of the amino acids included correspond to reference [17] and not to those given in the crystal data
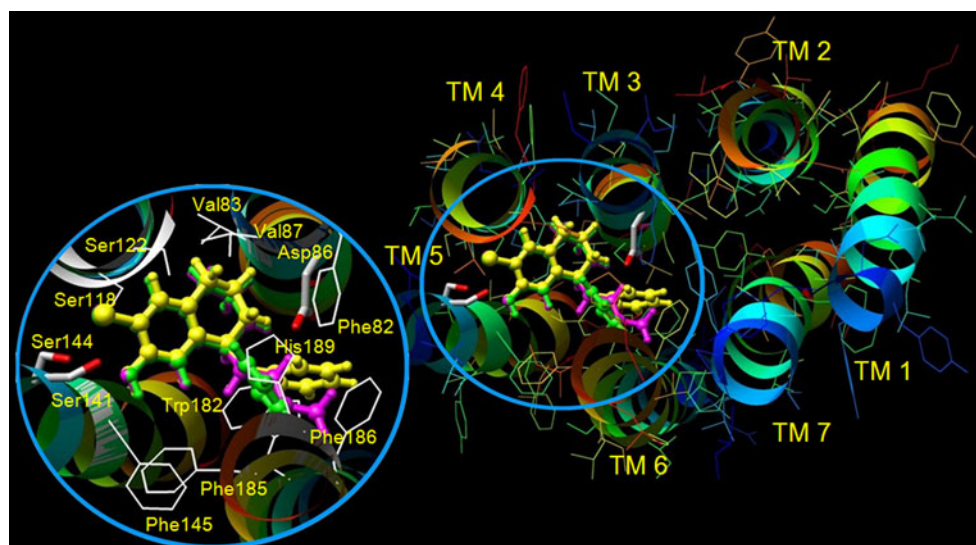
simulation period. Similar results were obtained for compounds **6**, **10**, **13** and **16**. However, for the rest of the BTHIQs evaluated here, such interactions were slightly weaker. It should be noted that in compounds **3**, **6**, **9**, **10**, **13** and **16**, the hydroxyl group on the ring-A is acting as a proton-donor; whereas the oxygen atom of the OH group of Ser141 is the proton-acceptor counterpart. In contrast, in the case of compounds **1**, **2**, **4**, **5**, **7**, **8**, **11**, **12**, **14** and **15**, the OH group of Ser 141 is the proton-donor and the methoxyl group on the ring-A is the acceptor counterpart. MD simulations predict that these interactions are weaker in comparison to those observed for hydroxyl ligands on the ring-A.

Aromatic side chains are bulky, have low barriers for rotation, and are ideal for adjusting to the changing conformation of the hydrophobic moiety of the ligand. In
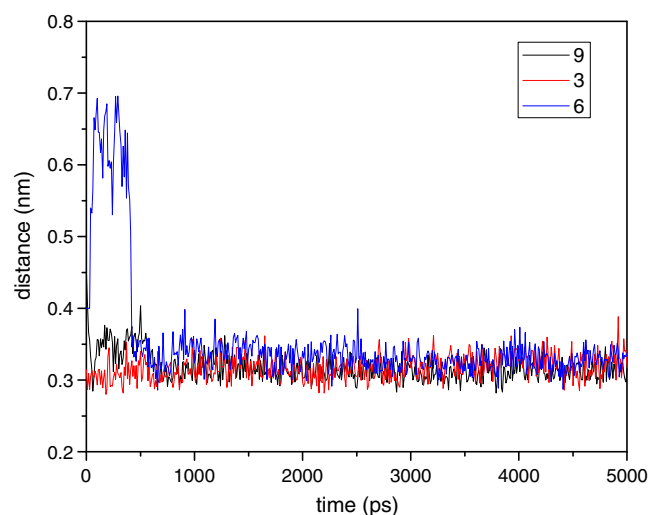
Fig. 4 Bond lengths obtained for the salt bridge between Asp 86 and the protonated amine group in compounds **3**, **6** and **9**
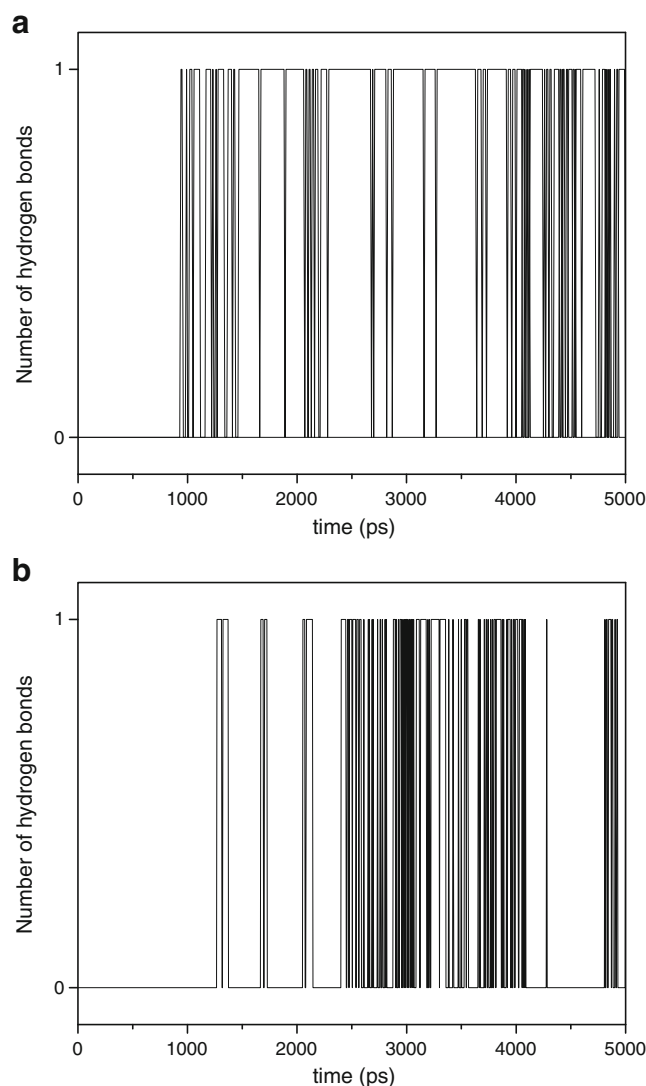
Fig. 5 Hydrogen bonds obtained for compounds **9** (**a**) and **3** (**b**). These interactions are between Ser 141 and the catecholic hydroxyls

the dopamine D2 receptor, the binding site proved to be aligned with aromatic side chains, and such residues can adjust to the different shapes and flexibility of the ligands in the binding site. Thus, Phe 82, Val 83 and Val 87(TM3); Phe 145 (TM5); and Trp 182, Phe 185, Phe 186 and His 189 (TM6) form a mostly hydrophobic pocket for ligands (Fig. 3).

It is interesting to note that the only structural differences between compounds **3**, **6**, **9**, **13** and **16** are the different substituents at C-1. Whereas compound **6** has a relatively rigidly held phenyl ring, the corresponding butyl, benzyl and halogenated-benzyl substituents on compounds **3**, **9**, **13** and **16**, respectively, are free to rotate, allowing better accommodation of these hydrophobic moieties to interact with the cluster of aromatic and non polar residues. These results might be better appreciated by observing the different conformational behaviors obtained for the torsional angles of their respective hydrophobic portions during the simulations (Figs. 5–8). The conformational behaviors observed for the torsional angles θ_1 and θ_2 of compound **9** are shown in Fig. 4. Whereas θ_1 is maintained relatively fixed at about 250° during the simulation (Fig. 6a), the torsional angle θ_2 displayed a high molecular flexibility, adopting conformations from 20° to 300° (Fig. 6b). Closely related results were obtained for the torsional angles θ_1 and θ_2 of compounds **13** and **16**. The hydrophobic portion of compound **3**, the butyl moiety, also displayed a high molecular flexibility. Figure 6 gives the conformational behaviors of torsional angles ϕ_1–ϕ_3 of compound **3**. The torsional angle ϕ_1 adopts a relatively rigid planar form close to 170° (Fig. 7a) but the other two torsional angles ϕ_2 and ϕ_3 displayed a high molecular flexibility (Fig. 7b and c, respectively). Very similar results were obtained for the butyl portion of compound **10**. In contrast, the conformational behavior obtained for the phenyl ring of compounds **4**–**6** displayed a very restricted molecular flexibility, keeping a spatial ordering almost perpendicular with respect to the rest of the molecule during the entire simulation (Fig. 8). The different affinities previously reported for compounds **6** and **9** suggest that the orientation of the substituent at C-1 may be a more important factor in the different effects on receptor affinity for the two ligands. This argument also applies to **3**, **10**, **13** and **16**, where the orientations of the butyl and halogenated-benzyl substituents are more favorable for hydrophobic interactions. Thus, the different affinities and selectivities obtained for these compounds might be explained, at least in part, by the different spatial orientations adopted by the varied hydrophobic portions located at C-1, which give different molecular interactions with the D2 receptor. These aspects are discussed in detail in terms of quantum mechanics calculations in the next section.

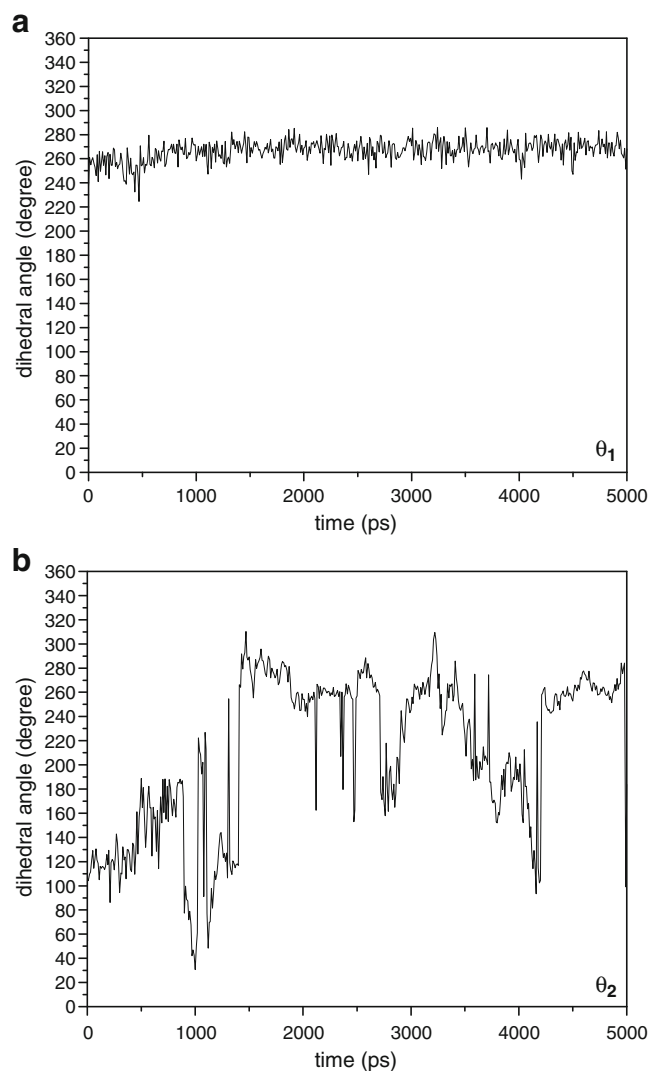In the next step of our study, we evaluated the binding energies (BE) obtained for the different com-

Fig. 6 Evolution of the angles θ_1 (**a**) and θ_2 (**b**) of compound **9** with time during the simulation

plexes. From the binding energies obtained in our MD simulations, one can distinguish a very good binder from a very weak binder ($-441,217.42$ kJ mol^{-1} for compound **9** vs $-441,015.35$ kJ mol^{-1} for compound **4**) but cannot distinguish ligands with similar binding affinities ($-441,217.42$ kJ mol^{-1} for compound **9** vs $-441,004.26$ kJ mol^{-1} for compound **3** and $-441,015.35$ kJ mol^{-1} for compound **4** vs $-440,155.71$ kJ mol^{-1} for compound **1** among other examples). This is not an unexpected result; can we realistically expect to make accurate and reliable predictions with what are decidedly crude representations of the molecular interactions involved in the binding process? Any model that neglects or only poorly approximates the terms that are playing determinant roles, such as, e.g., lone pair directionality in hydrogen bonds, explicit π-stacking polarization effects, hydrogen bonding networks, induced fit, and conformational entropy,
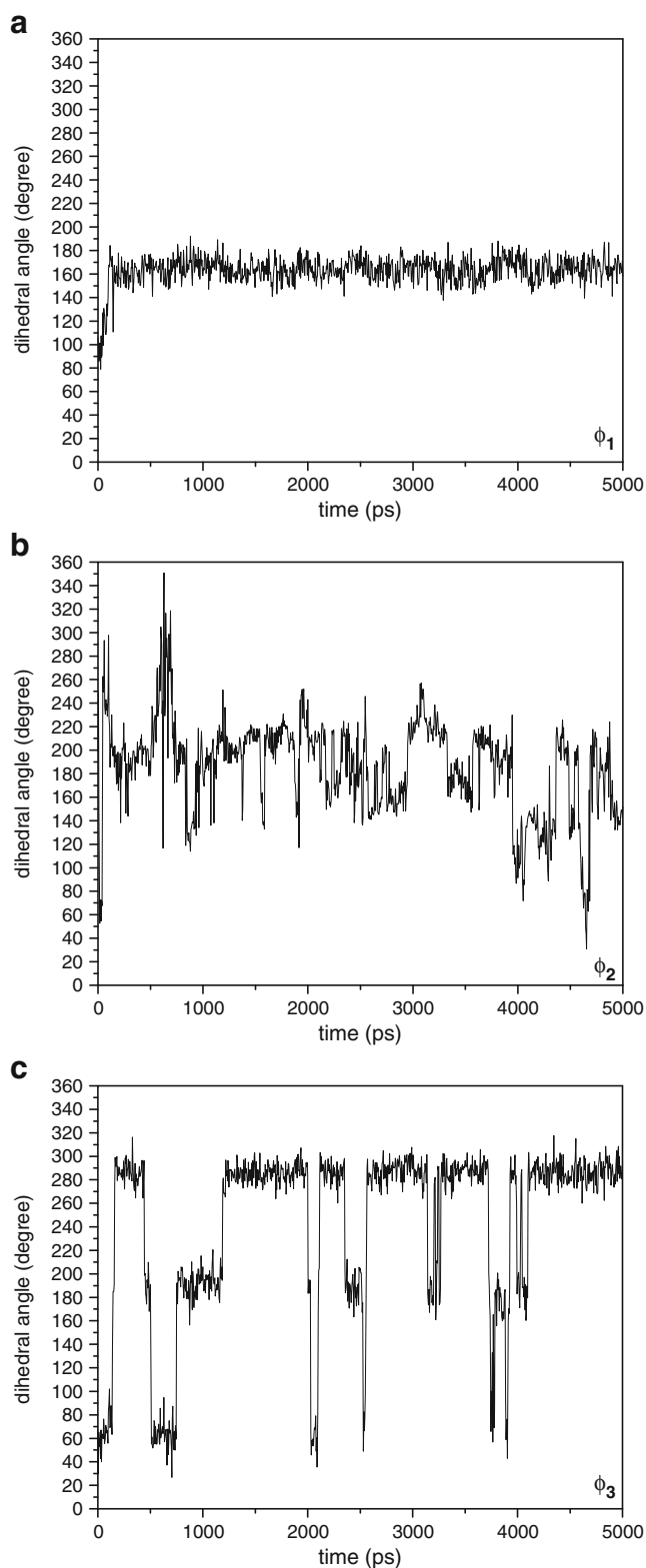
a

b

c

Fig. 7 Evolution of the angles ϕ_1 (**a**), ϕ_2 (**b**) and ϕ_3 (**c**) of compound **3** with time during the simulation
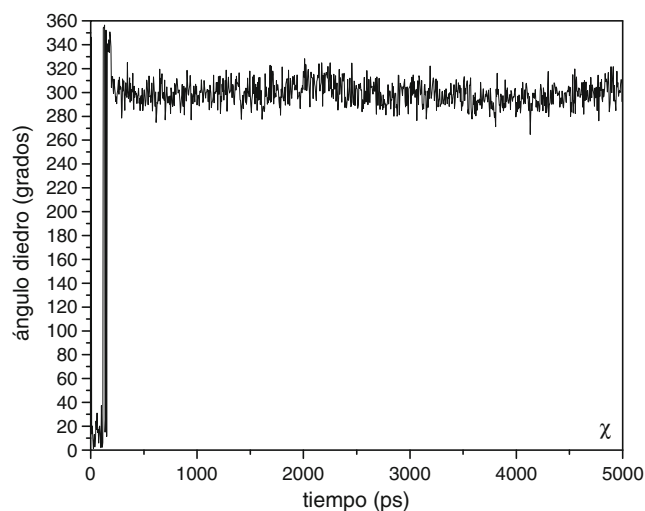

Fig. 8 Evolution of the angle χ of compound **6** with time during the simulation

energies. There are several works supporting this concept in the literature [50, 51].

At this stage of our work, we considered the trend predicted for the MD simulations as certainly significant but, on the other hand, we might be reluctant to assign it a quantitative significance, because of the approximations involved in this mode of approach. It should be noted that we are dealing with relatively weak interactions and therefore MD simulations might underestimate such interactions. Thus, in the next step, we optimized reduced model systems using combined semiempirical, ab initio and DFT calculations.

Quantum mechanics calculations

AM1 calculations combined with RHF/6-31G(d) and DFT [B3LYP/6-31G(d,p)] optimizations were performed by considering all receptor amino acids that could interact after initial positioning of the ligands against Asp 86 and Ser 141 residues. The binding pocket designed in this way (Fig. 3) provided data that matched experimental results previously reported from binding assays [8, 9].

Figure 9a shows ligand **13** interactions with the D2 DR optimized using quantum mechanical calculations. The salt bridge between the protonated amino group and the carboxyl group of Asp 86, as well as the hydrogen bond between the 7-hydroxyl group with Ser 141 can be seen in this figure. From Fig. 9a it is clear that a strong salt bridge exists in this compound between the protonated amino groups and the carboxyl group of Asp 86 (calculated distance of 3.47Å). The hydrogen bond between **13** and Ser 141 is a bifurcated interaction in which the oxygen atom of the hydroxyl group and the oxygen of carbonyl group of Ser 141 are the proton-acceptors, giving interatomic

among others, cannot reasonably be expected to distinguish between compounds possessing relatively similar binding

Fig. 9 Interactions of compound **13** (**a**) and **11** (**b**) with the binding pocket of D2 DR. Spatial view of two interactions: salt bridge (Asp 86 with protonated amino group) to the *right* and hydrogen bond between meta-hydroxyl group with Ser 141to the *left*
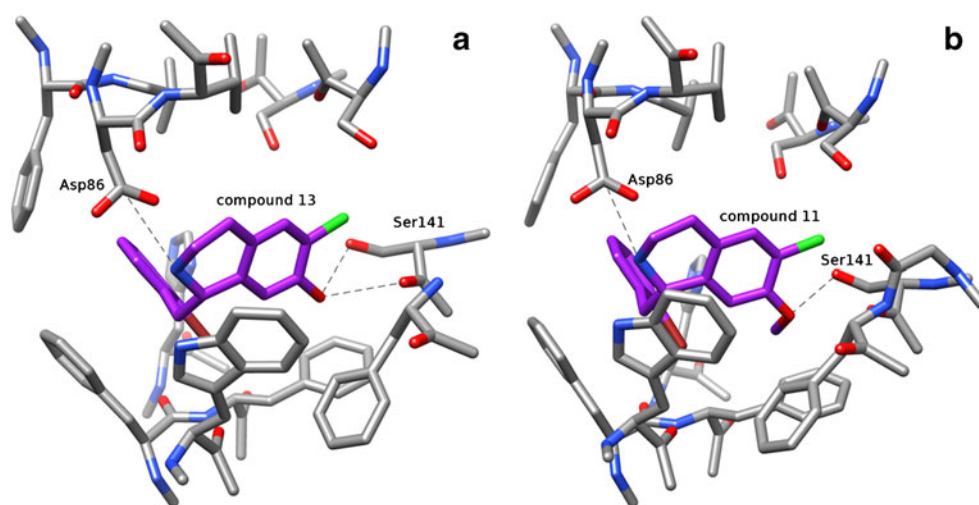
distances of 2.28 Å and 2.40 Å, respectively. Figure 9b shows ligand **11** interaction with the D2 DR. In this case, the 7-methoxyl group acts as proton-acceptor while the hydroxyl group of Ser 141 is the proton-donor, displaying an interatomic distance of 2.32 Å.

Table 1 gives the BE calculated for the different complexes using RHF/6-31 G(d) and B3LYP/ 6-31 G(d,p)

calculations. All compounds possessing 7-methoxyl groups displayed higher BE with respect to the 7- hydroxyl homologues (cf. **1** with **3**; **4** with **6**; **7** with **9**; **11** with **13**, and **14** with **16**). Previously, we reported that a 7-hydroxyl group acting as a proton-donor gives a stronger hydrogen bond than those derivatives possessing a 7-methoxyl group [16]. The present results are in agreement with previously
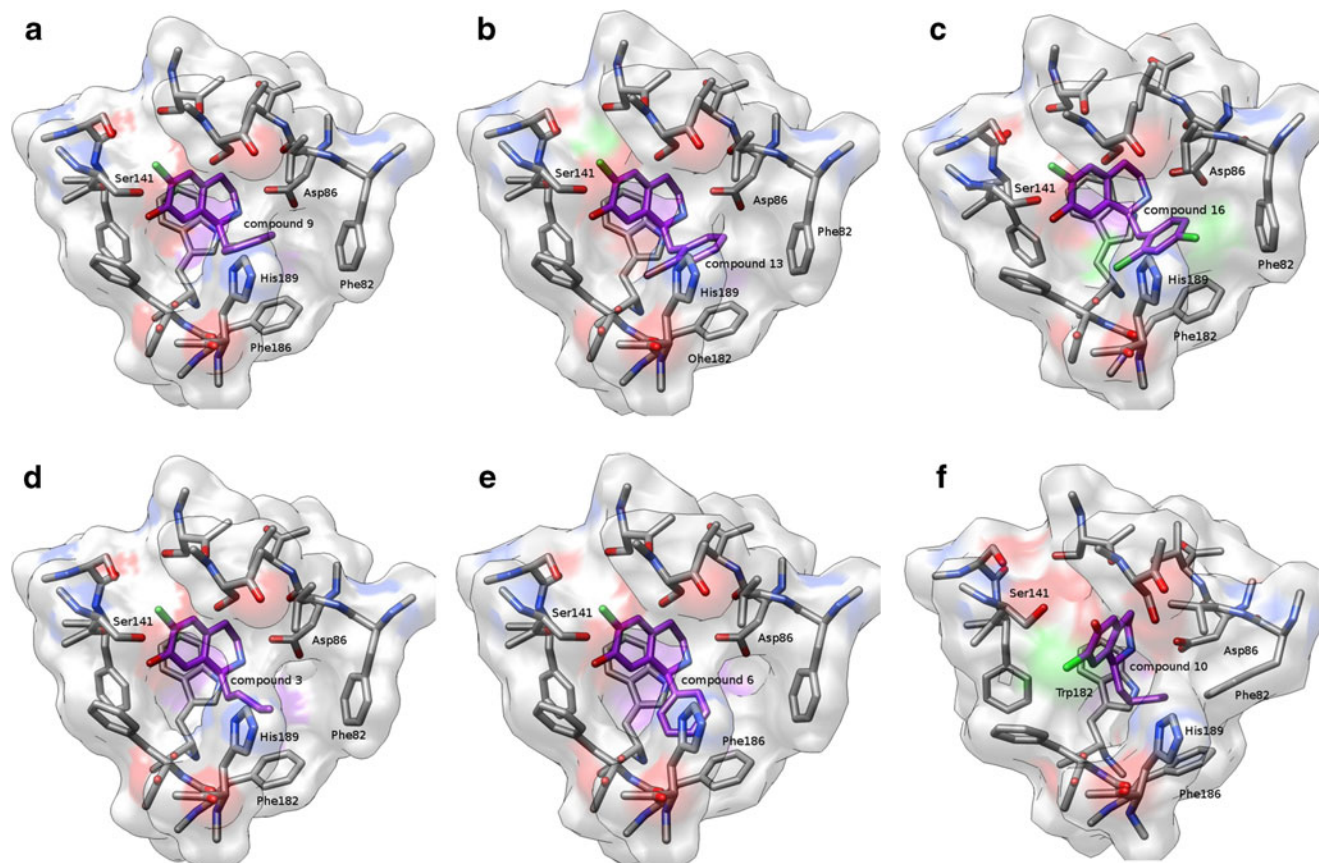

Fig. 10 Interactions of compound **9** (**a**), **13** (**b**), **16** (**c**), **3** (**d**), **6** (**e**) and **10** (**f**) with the binding pocket D2 DR. Different spatial views show the hydrophobic interactions at the hydrophobic zone

reported calculations for isolated and solvated molecules, as well as with previously reported experimental binding affinities [8, 9] (see Table 1).

Figure 10a shows ligand **9** interactions with the binding pocket. In this case, a different spatial view with respect to Fig. 9 is shown in order to better appreciate the hydrophobic interactions. From this figure, we can observe that the benzyl group of **9** adopts an adequate conformation to interact with Phe 186, Phe 82 and His 189. A similar spatial ordering was obtained for compounds **13** and **16** (Fig. 10b and c, respectively). For compounds **13** and **16**, the halogen substituent confers a higher polarizability on the benzyl group, allowing a stronger hydrophobic interaction. In should be noted that compound **13** displayed the highest of p*K*i value in this series. These hydrophobic interactions could explain, at least in part, the strongest affinity obtained for this compound. The butyl group of **3** displays a spatial ordering closely related to that of the benzyl group of **9**, **13** and **16**, also giving closely related hydrophobic interactions with the same hydrophobic residues (Fig. 10d). In contrast, the phenyl group of **6** displayed a different spatial ordering, giving adequate distance to interact only with Phe 186 (Fig. 10e). Interestingly, the bonding energies obtained for these complexes are: **13**/D2 DR < **10**/D2 DR < **9**/D2 DR < **3**/D2 DR < 1**6**/D2 DR, which are in complete agreement with their respective p*K*i values obtained from our previous experimental results (see Table 1). Compound **10** adopts a different spatial ordering at the binding site; thus, the butyl portion of this compound interacts with three aromatic residues: Trp 182, Phe 82 and Phe 186 (Fig. 10f). Compounds **4**–**6** possess a phenyl ring perpendicular to the rest of the ligand from the ring containing the protonated nitrogen [52]. These compounds docked in the D2 receptor model have few interactions in the binding pocket because their 1-phenyl substituents extend toward the extracellular surface of the receptor, parallel to the helix axes. These results are in agreement with those previously reported [23]. Thus, it appears that the shape and flexibility of the side chain at the C-1 position affects the receptor subtype selectivity of ligands to an extent that depends on the geometry, flexibility and stacking potential of ligand substituents. Lan et al. [23], reported that the D1 selective ligand SCH23390 contains a phenyl ring perpendicular to the rest of the molecule and the membrane plane, and parallel to the helix axes, which could explain its selectivity. Our results are in agreement with those results. Compounds type **4**–**6** in this series displayed a conformational behavior closely related to that reported for SCH23390.

Regarding the general structure of BTHIQs reported here, it is reasonable to think that the presence of a chlorine atom at C6, and consequently halogen bonding interactions, could be operative for the ligand–receptor complex formation. Thus, this chlorine possibly could be interacting

through either a positive sigma hole with a negative site in its vicinity or through its negative lateral ring of electrostatic potential with a positive site in the vicinity. A comprehensive study on electrostatically driven non-covalent interactions has been reported recently by Politzer et al. [53]. In this latter article, the possibility that halogen and other σ-hole interactions can be competitive with hydrogen bonding has been clearly established. Unfortunately, from the limited information obtained from our relatively low-level theory calculations, it is not possible to properly determine if the halogen bonding interactions could take place here. It is clear that further, more accurate calculations, as well as quantum atoms in molecules (QAIM) [54, 55] analysis are necessary for a detailed description of these interactions. Such calculations are now in progress in our laboratory and will be reported later in a separate paper.
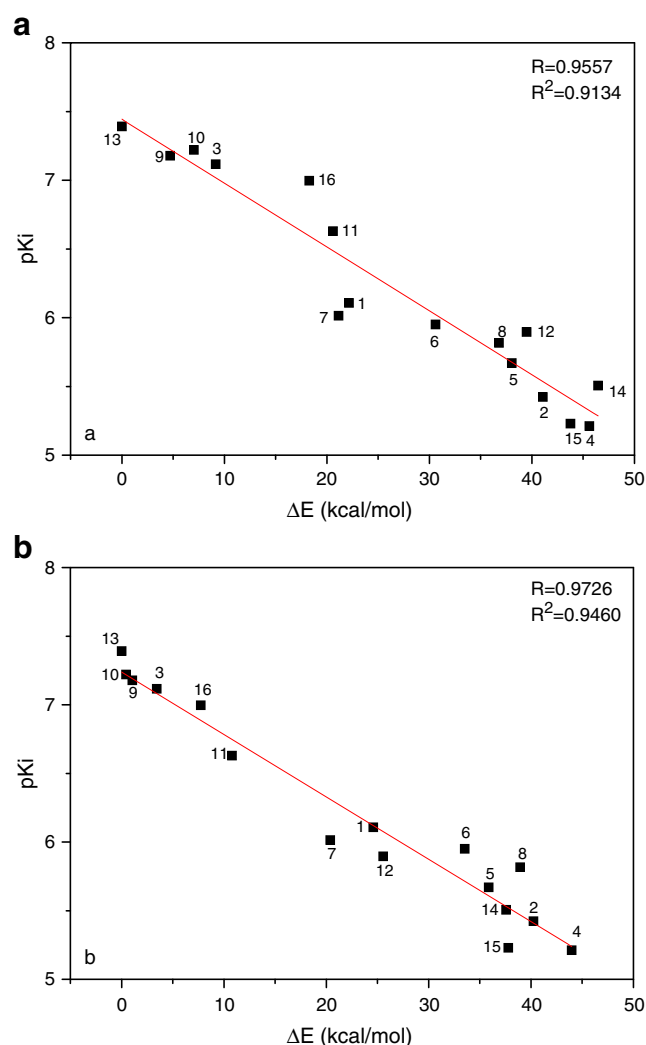

Fig. 11 Correlations obtained between the experimental p*K*i values versus the binding energies (BE) calculated from **a** ab initio [RHF/631G(d)] calculations, and **b** DFT (B3LYP/6-31G(d,p)) computations

Figure 11a gives a graphical representation of the calculated BEs obtained from RHF/6-31G(d) calculations versus experimental pKi values, obtained in binding studies in rat striatum [8, 9]. This figure has a correlation coefficient $R^2 = 0.9134$. Theisresult is very satisfactory when one considers the type of approximations used. Figure 11b shows the same correlations but in this case using BEs obtained from DFT [B3LYP/6-31G(d,p)] calculations. In this draft, the correlation coefficient is $R^2 = 0.9460$ indicating that DFT calculations give a better correlation with the experimental data. Although both linear correlations are good enough to predict the biological activity of BTHIQs, it is clear that DFT calculations give a significantly better correlation with respect to RHF computations. This is particularly evident from the high squares of correlation coefficients, r^2 obtained using RHF and DFT calculations (0.9134 and 0.9460, respectively). From our results, it is clear that the predicted first-principles structure of the primary binding pocket of D2 DR leads to correct predictions of the critical residues for binding THIQs, and gives relative binding affinities that correlate fairly well with those obtained in experiments performed in native tissue. This good correlation provides additional validation for the predicted structure and function.

It should be noted that the AM1 method it is not adequate to describe the hydrogen bonds. In addition, the ab initio and DFT calculations performed here probably do not properly consider the dispersion interactions. Fortunately, in this case it appears that such limitations are not severe enough to prevent us obtaining our objectives. Such an assumption appears to be reasonable, considering the significant correlation obtained between the experimental data and the theoretical calculations performed. However, we cannot exclude that a kind of error-cancellation could have taken place in this case. Thus, it must be pointed out that the approaches used in this study could be operative only for THIQs and structurally related compounds. To extend these approaches to other compounds possessing different structures would require additional validation and more accurate calculations.

Conclusions

A molecular modeling study on 16 BTHIQs acting as dopaminergic ligands was carried out. By combining MD simulations with ab initio and DFT calculations, a simple and generally applicable procedure to evaluate the binding energies of BTHIQs interacting with the D2 DR is reported here, providing a clear picture of the binding interactions of BTHIQs from both structural and energetic viewpoints. Thus, our results give interesting information that may be helpful in obtaining a better understanding of the molecular interactions between BTHIQs and the D2 DR.

A significant correlation between binding energies obtained from DFT calculations and experimental pKi values was obtained. These results could predict the potential dopaminergic effect of non-synthesized BTHIQs with an acceptable degree of accuracy. Such information could be essential in determining a priori the putative activity of new BTHIQ derivatives. It is prudent to remark that the excellent correlation obtained here between experimental data and the theoretical calculations performed here could be limited to BTHIQs and structurally related compounds. However, we believe our results may be helpful in the structural identification and understanding of the minimum structural requirements for these molecules, and can provide a guide to the design of BTHIQs with this biological activity.

References

1. Oloff S, Mailman RB, Tropsha A (2005) Application of validated QSAR models of D1 dopaminergic antagonist for database mining. J Med Chem 48:7322–7332
2. Sit SY, Xie K, Jacutin-Porte S, Boy KM, Seanz J, Taber MT, Gulwadi AG, Korpinen CD, Burris KD, Molski TF, Ryan E, Xu C, Verdoorn T, Johnson G, Nichols DE, Mailman RB (2004) Synthesis and SAR exploration of dinapsoline analogues. Bioorg Med Chem 12:715–734
3. Strange PG (1997) Dopamine receptors. Tocris Reviews 15. Tocris Cookson, Bristol
4. Zhang A, Neumeyer JL, Baldessarini RJ (2007) Recent progress in development of dopamine receptor subtype-selectiveagents: potential therapeutics for neurological and psychiatric disorders. Chem Rev 107:274–302
5. Protais P, Arbaoui J, Bakkali EH, Bermejo A, Cortes D (1995) Effects of various isoquinoline alkaloids on in vitro ^{3}H-dopamine uptake. J Nat Prod 58:1475–1484
6. Bermejo A, Protais P, Blázquez MA, Rao KS, Zafra-Polo MC, Cortes D (1995) Dopaminergic isoquinoline alkaloids from rotos of *Xylopia papuana*. Nat Prod Res 6:57–62
7. Cabedo N, Protais P, Cassels BK, Cortes D (1998) Synthesis and dopamine receptor selectivity of the benzyltetrahydroisoquinoline, (*R*)-(+)-*nor*-roefractine. J Nat Prod 61:709–712
8. Cabedo N, Andreu I, Ramírez de Arellano MC, Chagraoui A, Serrano A, Bermejo A, Protais P, Cortes D (2001) Enantioselective syntheses of dopaminergic (*R*)- and (*S*)-benzyltetrahydroisoquinolines. J Med Chem 44:1794–1801
9. Berenguer I, El Aouad N, Andujar S, Romero V, Suvire F, Freret T, Bermejo A, Ivorra MD, Enriz RD, Boulouard M, Cabedo N, Cortes D (2009) Tetrahydroisoquinolines as dopaminergic ligands: 1-butyl-7-chloro-6-hydroxy-tetrahydroisoquinoline, a new compound with antidepressant-like activity in mice. Bioorg Med Chem 17:4968–4980
10. El Aouad N, Berenguer I, Romero V, Marín P, Serrano A, Andujar S, Suvire F, Bermejo A, Ivorra MD, Enriz RD, Cabedo N, Cortes

D (2009) Structure–activity relationship of dopaminergic halogenated 1-benzyl-tetrahydroisoquinoline derivatives. Eur J Med Chem 44:4616–4621

11. Andreu I, Cabedo N, Torres G, Chagraoui A, Ramirez de Arellano MC, Gil S, Bermejo A, Valpuesta M, Portais P, Cortes D (2002) Syntheses of dopaminergic 1-cyclohexylmethyl-78-dioxygenated tetrahydroisoquinolines by selective heterogeneous tandem hydrogenation. Tetrahedron 58:10173–10179

12. Bermejo A, Andreu I, Suvire F, Léonce S, Caignard DH, Renard P, Pierré A, Enriz RD, Cortes D, Cabedo N (2002) Syntheses and antitumor targeting G1 phase of the cell cycle of benzoyldihydroisoquinoline. J Med Chem 45:5058–5068

13. Suvire FD, Andreu I, Bermejo A, Zamora MA, Cortes D, Enriz RD (2003) Conformational study of N-alkyl-benzyltetrahydroisoquinolines alkaloid. J Mol Struct THEOCHEM 666–667:109–116

14. Suvire FD, Cabedo N, Chagraoui A, Zamora MA, Cortes D, Enriz RD (2003) Molecular recognition and binding mechanism of N-alkyl-benzyltetrahydro-isoquinolines to the D_1 dopamine receptor. A computational approach. J Mol Struc THEOCHEM 666–667:455–467

15. Andreu I, Cortes D, Protais P, Cassels BK, Chagraoui A, Cabedo N (2000) Preparation of dopaminergic N-Alkyl-benzyltetrahydroisoquinolines using a 'One-Pot' procedure in acid medium. Bioorg Med Chem 8:889–895

16. Doi S, Shirai N, Sato Y (1997) Abnormal products in the Bischler-Napieralski isoquinoline synthesis. J Chem Soc Perkin Trans 1:2217–2221

17. Andujar SA, Migliore de Angel BM, Charris JE, Israel A, Suárez-Roca H, López SE, Garrido MR, Cabrera EV, Visual G, Rosales C, Suvire FD, Enriz RD, Angel-Guío JE (2008) Synthesis dopaminergic profile and molecular dynamics calculations of N-aralkyl substituted 2-aminoindans. Bioorg Med Chem 16:3233–3244

18. Kalani MYS, Vaidehi N, Hall SE, Trabanino RJ, Freddolino PL, Kalani MA, Floriano WB, Wai Tak Kam V, Goddard WA III (2005) The predicted 3D structure of the human D2 dopamine receptor and the binding site and binding affinities for agonists and antagonists. Proc Natl Acad Sci USA 101:3815–3820

19. Becker OM, Marantz Y, Shacham S, Inbal B, Heifetz A, Kalid O, Bar-Haim S, Warshaviak D, Fichman M, Noiman SG (2004) Protein-coupled receptors: in silico drug discovery in 3D. Proc Natl Acad Sci USA 101:11304–11309

20. Micheli F, Bonanomi G, Blaney FE, Braggio S, Capelli AM, Checchia A, Curcuruto O, Damiani F, Di Fabio R, Donati D, Gentile G, Gribble A, Hamprecht D, Tedesco G, Terreni S, Tarsi L, Lightfoot A, Stemp G, MacDonald G, Smith A, Pecoraro M, Petrone M, Perini O, Piner J, Rossi T, Worby A, Pilla M, Valerio E, Griffante C, Mugnaini M, Wood M, Scott C, Andreoli M, Lacroix L, Schwarz A, Gozzi A, Bifone A, Ashby CR Jr, Hagan JJ, Heidbreder C (2007) 124-Triazol-3-yl-thiopropyl-tetrahydrobenzazepines: a series of potent and selective dopamine D3 receptor antagonists. J Med Chem 50:5076–5089

21. Ribeiro AA, Horta BAC, de Alencastro RB (2008) MKTOP: a program for automatic construction of molecular topologies. J Braz Chem Soc 19:1433–1435

22. Manzour A, Meng F, Meador-Woodruff JH, Taylor LP, Civelli O, Akil H (1992) Site-directed mutagenesis of the human dopamine D_2 receptor. Eur J Pharm Mol Pharmacol Sec 227:205–214

23. Lan H, DuRand CJ, Teeter MM, Neve KA (2006) Structural determinants of pharmacological specificity between D_1 and D_2 dopamine receptors. Mol Pharmacol 69:185–194

24. Berendsen HJC, Van der Spoel D, Van Drunen R (1995) GROMACS: a message-passing parallel molecular dynamics implementations. Comput Phys Commun 91:43–56

25. Lindahl E, Hess B, van der Spoel D (2001) GROMACS 3.0: a package for molecular simulations and trajectory analysis. J Mol Model 7:306–317

26. van Buuren AR, Marrink SJ, Berendsen HJC (1993) A molecular dynamics study of the decane/water interface. J Phys Chem 36:9206–9212

27. Mark AE, van Helden SP, Smith PE, Janssen LHM, van Gunsteren WF (1994) Convergence properties of free energy calculations. A-cyclodextrin complexes as a case study. J Am Chem Soc 116:6293–6302

28. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935

29. van Buuren AR, Berendsen HJC (1993) Molecular dynamics simulation of the stability of a 22 residue alpha-helix in water and 30% trifluoroethanol. Biopolymers 33:1159–1166

30. Liu H, Muller-Plathe F, van Gunsteren WF (1995) A force field for liquid dimethyl sulfoxide and physical properties of liquid dimethyl sulfoxide calculated using molecular dynamics simulation. J Am Chem Soc 117:4363–4366

31. Miyamoto S, Kollman PA (1992) SETTLE—an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 13:952–962

32. Berendsen HJC, Postma HJC, Van Gunsteren WF, Hermans WF (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular forces. Reidel, Dordrecht, pp 331–342

33. 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

34. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593

35. Luty B, Tironi IG, van Gunsteren WF (1995) Lattice–sum methods for calculating electrostatic interactions in molecular simulations. J Chem Phys 103:3014–3021

36. Zimmerman K (1991) All purpose molecular mechanics simulator and energy minimizer. J Comput Chem 12:310–319

37. Ferguson DM (1995) Parameterization and evaluation of a flexible water model. J Comput Chem 16:501–511

38. Berendsen HJC, Postma JPM, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690

39. Hess B, Bekker H, Berendsen HJC, Fraaije JG (1997) E.M. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472

40. Kampmann T, Mueller DS, Mark AE, Young PR, Kobe B (2006) The role of histidine residues in low-pH-mediated viral membrane fusion. Structure 14:1481–1487

41. Teeter MM, Froimowitz MF, Stec B, Durand CJ (1994) Homology modeling of the dopamine D2 receptor and its testing by docking of agonists and tricyclic antagonists. J Med Chem 37:2874–2888

42. Neve KA, Cumbay MG, Thompson KR, Yang R, Buck DC, Watts VJ, Durand CJ, Teeter MM (2001) Modeling and mutational analysis of a putative sodium-binding pocket on the dopamine D_2 receptor. Mol Pharmacol 60:373–381

43. 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, 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 B.05. Gaussian Inc, Pittsburgh

44. 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

45. Hjerde E, Dahl SG, Sylte I (2005) Atypical and typical antipsychotic drug interactionswith the dopamine D_2 receptor. Eur J Med Chem 40:185–194

46. Cho W, Taylor LP, Mansour A, Akil A (1995) Hydrophobic residues of the D_2 dopamine receptor are important for binding and signal transduction. J Neurochem 65:2105–2115

47. Cox BA, Henningsen RA, Spanoyannis A, Neve RL, Neve KA (1992) Contributions of conserved serine residues to the interactions of ligands with dopamine D_2 receptors. J Neurochem 59:627–635

48. Wiens BL, Nelson CS, Neve KA (1998) Contribution of serine residues to constitutive and agonist-induced signaling via the D2Sdopamine receptor: evidence for multiple agonist-specific active conformations. Mol Pharmacol 54:435–444

49. Wilcox RE, Huang WH, Brusniak MYK, Wilcox DM, Pearlman RS, Teeter MM, Durand CJ, Wiens BL, Neve KA (2000) CoMFA-based prediction of agonist affinities at recombinant wild type versus serine to alanine point mutated D_2 dopamine receptors. J Med Chem 43:3005–3019

50. Page CS, Bates PA (2008) Can MM-PBSA calculations Predict the specifiicties of protein kinase inhibitors? J Comput Chem 27:1990–2007

51. Mertz KM (2010) Limits of free energy computation for protein–ligand interactions. J Chem Theor Comput 6:1769–1776

52. Breneman CM, Wiberg KB (1990) Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. J Comput Chem 11:361–373

53. Politzer P, Murray JS, Clark T (2010) Halogen bonding: an electrostatically-driven highly directional noncovalent interaction. Phys Chem Chem Phys 12:7748–7757

54. Bader RFW (1990) Atoms in molecules. A quantum theory. Oxford University Press, Oxford

55. Popelier PLA (1999) Atoms in molecules. An introduction. Pearson, Harlow

ORIGINAL PAPER

Triethanolammonium acetate as a multifunctional ionic liquid in the palladium-catalyzed green Heck reaction

**Zorica D. Petrović · Svetlana Marković ·
Vladimir P. Petrović · Dušica Simijonović**

Abstract An efficient green Heck reaction protocol was performed using a triethanolammonium acetate ionic liquid–palladium(II) catalytic system. The ionic liquid used acts as a reaction medium, base, precatalyst-precursor, and mobile support for the active Pd species. Our experimental investigation indicates that performing the Heck reaction in ionic liquid is superior to the same procedure carried out in triethanolamine. The mechanism of the reaction of triethanolammonium acetate with $PdCl_2$ was examined using density functional theory (M06 method). It was found that two Pd(II) complexes are formed, one of which acts further as a precatalyst yielding catalytically active Pd(0) complex. The calculated activation energies are in agreement with our experimental findings.

Keywords Triethanolamine ionic liquid · Green Heck reaction · Reaction mechanism · Pd(II) precatalyst · Pd(0) catalytically active complex

Introduction

The palladium-catalyzed arylation of olefins, known as the Heck reaction, is one of the most important methods of carbon–carbon bond formation in organic synthesis [1–4]. Due to its synthetic versatility, the Heck reaction attracts the attention of both experimental [5–9] and theoretical chemists [10–22]. Much effort has been devoted to elucidation of the Heck reaction mechanism [10–12], including catalyst preactivation [13, 14], oxidative addition [15–21], and reductive elimination [22, 23].

In the last decade, phosphine-free Pd complexes have been introduced as a less complicated and environmentally more desirable alternative to the original Pd-phosphine catalysts [1]. In addition, the application of ionic liquids has been put forward as a very useful substitute for hazardous and volatile polluting organic solvents [24–27]. One of the predominant applications of ionic liquids focuses on homogeneous catalysis [24, 28]. Nowadays, multifunctional ionic liquids that can serve as good coordinating ligands, green solvents, as well as recyclable and mobile supports for palladium catalyst systems, are used. Functional groups that can complex palladium, such as amine, amide, nitrile, ether, alcohol, urea or thiourea [29, 30] have been introduced into the cationic and/or anionic moiety of ionic liquids.

Here, we report the use of triethanolammonium acetate [TEA][HOAc] as a multifunctional ionic liquid in a green Heck reaction. The investigated ionic liquid acts as a green solvent, base, precatalyst-precursor, and mobile support for the active Pd species. Our work focused on testing the efficiency of the obtained ionic liquid–palladium catalytic system in the Heck reaction, and on a density functional theory (DFT) investigation of the possible mechanism of in situ formation of the Pd(II) precatalyst. Since very little is known about the mechanism of formation and molecular structure of Pd(0) complexes [13, 14], which are generally accepted as the catalytically active forms, our additional goal was to elucidate the mechanism of reduction of the Pd(II) precatalyst to the catalytically active Pd(0) form.

Z. D. Petrović · S. Marković (✉) · V. P. Petrović · D. Simijonović
Faculty of Science, University of Kragujevac,
12 Radoja Domanovića,
34000 Kragujevac, Serbia
e-mail: mark@kg.ac.rs

Computational methods

The geometrical parameters of all stationary points and transition states were optimized in vacuum, using Gaussian09 [31]. All calculations were performed using the M06 functional [32]. This hybrid meta functional is a functional with good accuracy "across-the-board" for transition metals, main group thermochemistry, medium-range correlation energy, and barrier heights [32]. The triple split valence basis set 6-311 G(d,p) was used for C, H, O, N, and Cl, whereas LANL2DZ+ECP [33] was employed for the Pd center. All calculated structures were confirmed to be local minima (all positive vibrational frequencies) for ground state structures, or first-order saddle points (one negative vibrational frequency) for transition state structures, by frequency calculations. The intrinsic reaction coordinates (IRCs), from the transition states down to the two lower energy structures, were traced using the IRC routine in Gaussian in order to verify that each saddle point is linked with two putative minima. Natural bond orbital analysis (Gaussian NBO version) was performed for all structures.

Results and discussion

The structure of ionic liquid triethanolammonium acetate ([TEA][HOAc]) was examined using DFT (Fig. 1). Our investigation revealed that there are no distinguishing cations and anions in [TEA][HOAc]. Bond distances reveal

strong hydrogen bonding in the ionic liquid. The NBO analysis was in accord with this finding, namely there is strong donation of density from the p orbital on N to the σ* antibonding O–H orbital of acetic acid. Owing to hydrogen bonding, it is obvious that molecular association is present. In addition, the molecule is polar (nitrogen and oxygen bear partial negative charge, whereas H1 is partially positively charged). These properties are characteristic of molecular liquids, and generally for ionic liquids. The role of the ionic liquid used and its triethanolamine Pd(II) complexes (Fig. 1) in a green Heck reaction protocol was investigated in this work.

Bearing in mind that the trans-[PdCl$_2$(DEA)$_2$] complex acts as a precatalyst in the Heck reaction in [DEA][HOAc] as reaction medium [34], we assumed that [TEA][HOAc] would be good reaction medium, and that the triethanolamine Pd(II) complex would also yield a catalytically active species in the Heck reaction. Thus, we decided to investigate the efficiency of the ionic liquid–palladium ([TEA][HOAc]/ Pd(II) complex) catalytic system. Above all we were interested in the mechanisms of in situ formation of the Pd(II) precatalyst and the catalytically active Pd(0) complex.

The Heck reaction was performed with the corresponding aryl halide (iodobenzene and bromobenzene) and activated olefine (methyl, ethyl, and butyl acrylate) in the investigated ionic liquid at 110°C, and with 2 mol% PdCl$_2$. A 100% conversion of acrylates to the trans cross-coupling products was achieved, with 99% selectivity. The yield of products

Fig. 1 Optimized geometries of triethanolammonium acetate ([TEA][HOAc]), and trans-[PdCl$_2$(TEA)$_2$] (Pd-N,N and Pd-N,O) complexes)
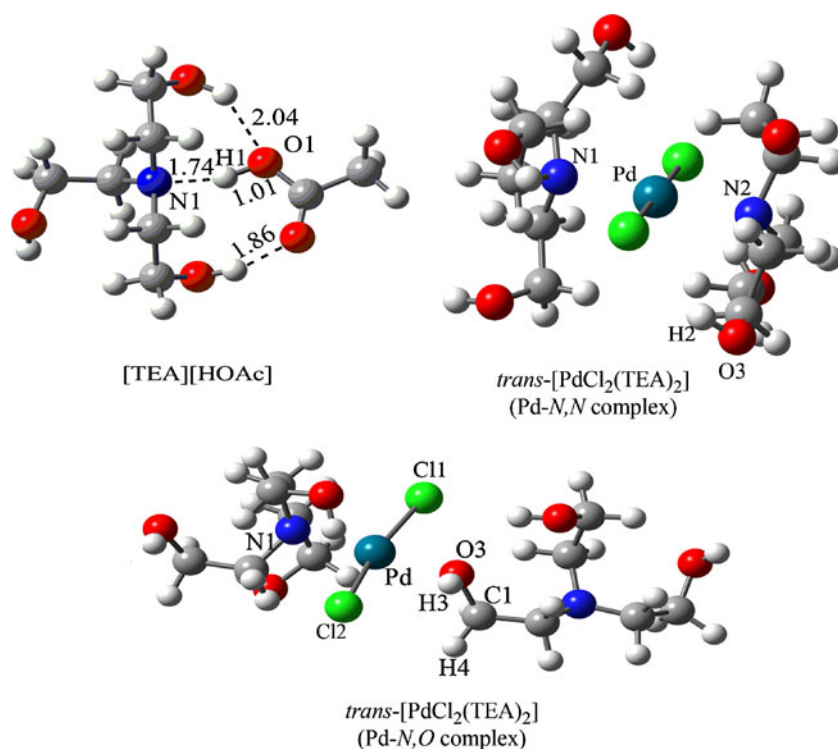
obtained in this reaction was very good (90–93%, Table 1). Next, we wished to compare the experimental procedure and yield of the ionic liquid–Pd(II) catalyzed reaction with the same reaction performed in the absence of ionic liquid, where TEA was used as a reaction medium (Table 1).

When the Heck reaction was performed in TEA, the yield was lower (75–81%). The lower yield of these reactions can be attributed to a competitive polymerization reaction. Our results indicate that the Heck reaction performed in the ionic liquid is superior compared to the procedure carried out in TEA. The ionic liquid used facilitates solubility of $PdCl_2$ and the appropriate Pd(II) catalyst precursor. It seems that the nature of the ionic liquid used increases the stability of the palladium catalyst, extending its lifetime. In addition, this catalytic system remains unchanged during the reaction. The reaction products can be separated via simple extraction, and ionic liquid–palladium catalytic system can be recycled easily.

After extraction of the products from the reaction mixture, the ionic liquid–palladium catalytic system was recovered and fresh reactants were charged. The ionic liquid–Pd catalytic system was recycled three times without significant loss in activity (yields of coupling products were up to 5% lower).

Mechanism of formation of Pd(II) complexes

In order to elucidate the mechanism of formation and structure of the Pd(II) precatalyst and catalytically active Pd (0) species, we investigated the possible mechanism of reaction of [TEA][HOAc] with $PdCl_2$. It turned out that two complexes can be formed in this reaction: Pd-N,N and Pd-N, O complexes (Figs. 1, 2). The selected bond distances of transition states and intermediates are given in Table 2, whereas the relative total energies, enthalpies, and free energies of all relevant species are provided in Table S1 in the Electronic Supplementary Material.

NBO analysis of $PdCl_2$ reveals that palladium is electron deficient, as it bears only four lone pairs in the d orbitals. On the other hand, the HOMO map of [TEA][HOAc] (Fig. S3) delineates the area in the molecule that is most electron sufficient (N1). In addition, the NBO charges of Pd (0.750), and N1 in [TEA][HOAc] (−0.582) indicate that palladium will perform an electrophilic attack on the nitrogen of the ionic liquid.

Our assumption was confirmed by revealing transition state TS1 (Fig. 2), which requires an activation energy of 127.0 kJ mol^{-1} (Table S1). In TS1, the Pd–N1 bond is being formed, whereas the hydrogen N1–H1 bond is being broken (Table 2). The O1–H1 bond is completely formed, implying that the acetic acid molecule leaves the reaction system as a solvent molecule. In this way, the tricoordinated intermediate I1 is formed (Fig. 2). Tricoordinated Pd complexes are known from the literature [12, 18–21]. NBO analysis of the intermediate shows that the p orbitals of the ligating atoms participate with over 80% in the bonds around palladium. Each Pd–Cl bond delocalizes into the adjacent σ* antibonding Pd–Cl orbital, whereas the Pd–N1 bond delocalizes into both σ* antibonding Pd–Cl orbitals. Palladium still bears only four lone electron pairs and, thus, it is electron deficient. The LUMO map of the intermediate (Fig. S4) is in accord with NBO analysis, and indicates Pd as a possible electrophilic site.

Table 1 Palladium-catalyzed Heck reaction in different reaction media. *[TEA][HOAc]* Triethanolammonium acetate, *TEA* triethanolamine

$$ArX + CH_2=CHR \xrightarrow[\text{[TEA][HOAc] or TEA}]{\text{2 mol.\% } PdCl_2,\ 110\ ^0C} ArCH=CHR$$

Entry	ArX	R	Reaction medium	Time (h)	Yield (%)[a]
1	C_6H_5I	$COOCH_3$	[TEA][HOAc]	12	93
2	C_6H_5Br	$COOCH_3$	[TEA][HOAc]	14	90
3	C_6H_5I	$COOC_2H_5$	[TEA][HOAc]	12	93
4	C_6H_5Br	$COOC_2H_5$	[TEA][HOAc]	14	91
5	C_6H_5I	$COOC_4H_9$	[TEA][HOAc]	12	93
6	C_6H_5Br	$COOC_4H_9$	[TEA][HOAc]	14	90
7	C_6H_5I	$COOCH_3$	TEA	12	81
8	C_6H_5Br	$COOCH_3$	TEA	14	77
9	C_6H_5I	$COOC_2H_5$	TEA	12	80
10	C_6H_5Br	$COO\ C_2H_5$	TEA	14	75
11	C_6H_5I	$COOC_4H_9$	TEA	12	80
12	C_6H_5Br	$COOC_4H_9$	TEA	14	75

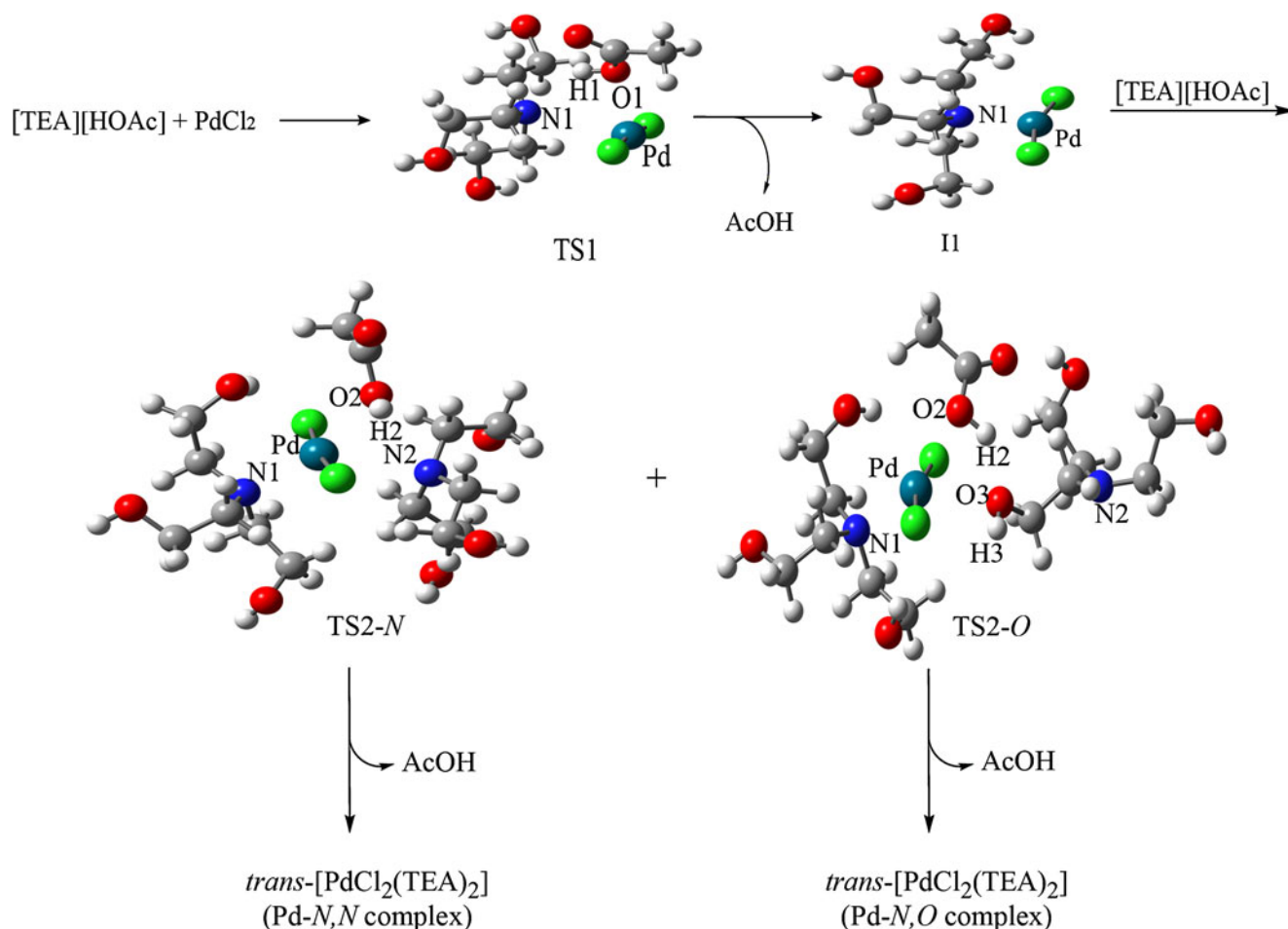[a] Isolated yield; only the *trans* product was detected by ^{1}H NMR

Fig. 2 Proposed mechanism for the formation of Pd-N,N and Pd-N,O complexes

Our investigations revealed two pathways for the electrophilic attack of Pd of I1 on another molecule of [TEA][HOAc]. The first pathway, in which Pd attacks nitrogen, occurs via transition state TS2-N (Fig. 2). An energy barrier of 139.7 kJ mol^{-1} is required for the formation of this transition state (Table S1). Similarly to the case of transition state TS1, the Pd–N2 bond is being formed, whereas the hydrogen N2–H2 bond is completely broken (Fig. 2, Table 2), again implying that acetic acid leaves the reaction system as a solvent molecule. In this way, a Pd-N,N complex is yielded (Figs. 1, 2). Our experiments confirm that the Pd–N,N complex is formed during the course of the Heck reaction, as it was identified in the reaction mixture by means of NMR spectroscopy. This identification was achieved by comparing the NMR spectrum of the reaction mixture to the NMR spectrum of the separately prepared Pd–N,N complex. Indeed, this complex was formed upon mixing of PdCl$_2$ with [TEA][HOAc] in molar ratio 1:2, at 100°C. The Pd-N,N complex exhibits a square planar coordination. According to NBO analysis, palladium is sp^{2}d hybridized, and builds covalent bonds with both nitrogens and both chlorines. The sp^3 orbitals of the Cl atoms and almost pure p orbitals of the N atoms participate with about 85% in the bonds around palladium. There is strong donation of density from each Pd–N bond to the adjacent σ* antibonding Pd–N orbital. As a consequence, the occupancies in the Pd–N orbitals are noticeable low (1.86).

The second pathway, in which Pd of I1 attacks the oxygen of [TEA][HOAc], proceeds via transition state TS2-O (Fig. 2), which has an activation barrier of 85.8 kJ mol^{-1} (Table S1). In TS2-O the Pd–O3 bond is being formed, whereas the hydrogen bond between N2 and H2 is completely broken (Fig. 2, Table 2). In this way the Pd–N,O complex is formed (Fig. 1). Our finding is supported by previously reported experimental results [35] that showed that oxygen coordination to Pd(II) and proton abstraction from the alcoholic group is possible, even in weak acidic medium. It is worth pointing out that acetate buffer is formed during our Heck reaction (AcOH/[TEA][HOAc], measured pH≈6).

Table 2 Selected bond distances (Å) in the investigated species for the mechanism of formation of Pd(II) complexes

Ionic liquid + PdCl$_2$	
N1-H1	1.05
O1-H1	1.61
Pd-Cl	2.31, 2.35
TS1	
N1-H1	2.11
O1-H1	0.98
Pd-N1	2.89
I1	
Pd-N1	2.08
Pd-Cl	2.32, 2.35
TS2-N	
Pd-N1	2.15
Pd-Cl	2.36, 2.38
N2-H2	3.40
O2-H2	0.98
Pd-N2	2.88
TS2-O	
Pd-N1	2.17
Pd-Cl	2.36, 2.38
N2-H2	4.42
O2-H2	0.98
Pd-N2	5.75
Pd-O3	2.59
Pd-N,N complex	
Pd-N1	2.17
Pd-Cl	2.38, 2.42
Pd-N2	2.17
Pd-N,O complex	
Pd-N1	2.13
Pd-N2	5.08
Pd-Cl	2.36, 2.40
Pd-O3	2.12

It should be emphasized that, in the Heck reaction with TEA, instead of [TEA][HOAc], a competitive polymerization reaction occurred, and formation of the Pd-N,N complex was not observed during the course of the reaction. In agreement with this experimental finding, our attempts to reveal a transition state for an electrophilic attack of Pd on the nitrogen or oxygen of TEA were unsuccessful.

Mechanism of formation of the Pd(0) complex

Analysis of the stability of the complexes obtained showed that the Pd–N,O complex is less stable than the Pd–N,N complex by 31.4 kJ mol^{-1}. This instability is a probable reason why it has not been identified in the reaction mixture. Thus we assume that it undergoes further transformation, yielding catalytically active Pd(0) complex.

The proposed mechanism is presented in Fig. 3. All relevant bond distances in transition states and intermediates are given in Table 3, whereas the relative total energies, enthalpies, and free energies of all relevant species are provided in Table S2.

NBO analysis of the Pd–N,O complex shows that Pd forms covalent bonds with both chlorines and nitrogen (Fig. 1). The p orbitals of the ligating atoms participate with over 80% in the bonds around palladium. The O3–H3 orbital and lone pair on O3 (sp^3 orbital) delocalize into formally empty p orbital (with little s and d mixing) on Pd, thus forming a coordinative Pd–O3 bond. Density is also donated from the sp^3 orbital of O3 to the σ* antibonding Pd–N1 orbital. The Pd–N1 bond delocalizes into both σ* antibonding Pd–Cl orbitals. Due to this, the occupancies of all palladium bonds are low (about 1.90). NBO analysis also reveals that the O3–H3 bond is particularly polar (NBO charges of O3 and H3 amount −0.697 and 0.517). Taking into account that O3 is engaged in the coordinative bond with Pd, and that H3 is acidic, we supposed that this H3 can be abstracted by the acetate buffer. Our assumption was confirmed by successful optimization of the intermediate I2 (Fig. 3). NBO analysis of I2 shows that p orbitals of the ligating atoms participate with over 80% in the bonds around palladium. Each Pd–Cl bond delocalizes into the adjacent σ* antibonding Pd-Cl orbital, whereas the Pd–O3 bond delocalizes into both σ* antibonding Pd–Cl orbitals. The lone pair on N1 delocalizes into σ* antibonding Pd–O3.

Our experience with the preactivation process of trans-[PdCl$_2$(DEA)$_2$] [13, 14] shows that β hydrogens (with respect to Pd) have affinity to coordinate to Pd. As H4 has the most favorable position, we supposed a nucleophilic attack of H4 to palladium(II) to be a plausible next step of the reaction. Our assumption was confirmed by revealing transition state TS3 (Fig. 3). The results of the IRC calculation for TS3 are presented in Fig. S5. The formation of TS3 has an energy barrier of 141.9 kJ mol^{-1} (Table S2). In TS3, Pd–O3 and C1–H4 bonds are being broken, while a Pd–H4 bond is being formed (Table 3). It is worth pointing out that hydrogen is transferred from carbon to palladium as a hydride ion. This transfer leads to the formation of the intermediate I3a (Fig. S2). The structure of I3a shows that it contains a completely separated molecule [N,N-bis(2-hydroxyethyl)amino-acetaldehyde] that probably acts further as a solvent molecule. For this reason we excluded the N,N-bis(2-hydroxyethyl)amino-acetaldehyde molecule (C$_6$H$_{13}$NO$_3$) from further consideration (I3 in Fig. 3). NBO analysis of I3 reveals that the intermediate complex exhibits a square planar coordination, where Pd forms covalent bonds with both chlorines and hydrogen. As for nitrogen, its almost pure p orbital donates density to the formally empty p orbital (with little s and d mixing) on Pd.
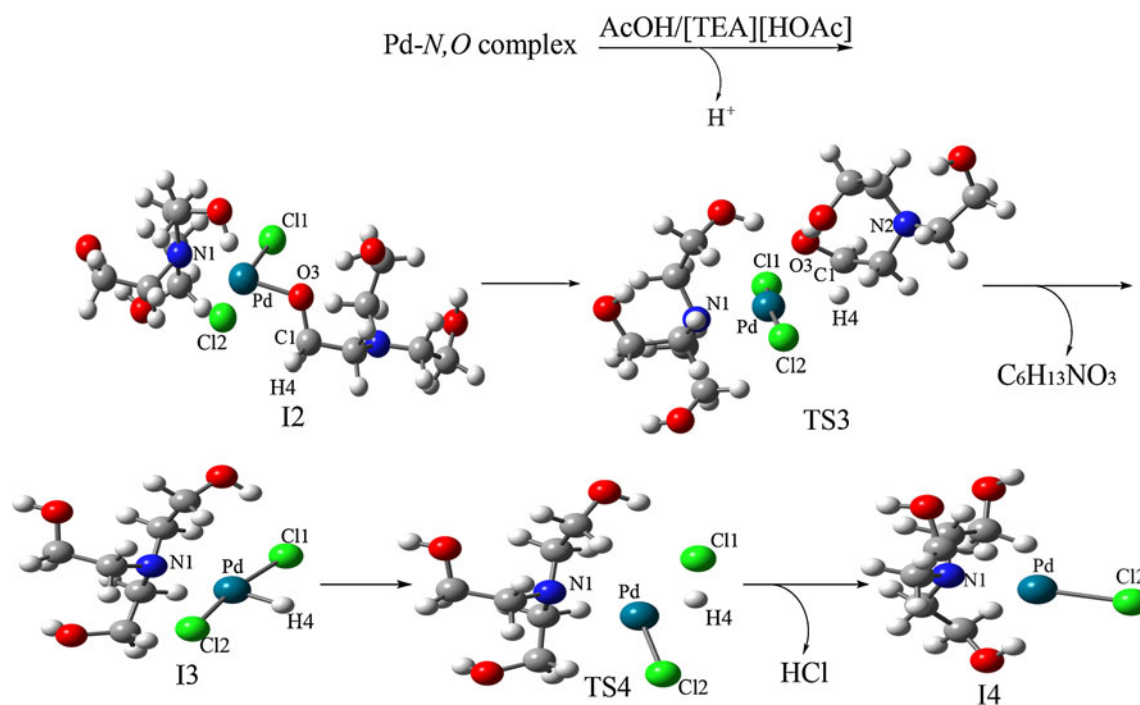
Fig. 3 Proposed mechanism for the formation of the Pd(0) complex

Table 3 Selected bond distances (Å) in the investigated species for the mechanism of formation of Pd(0) complex

I2	
Pd-N1	2.25
Pd-N2	4.87
Pd-Cl1	2.40
Pd-Cl2	2.43
Pd-O3	1.99
C1-H4	1.10
Pd-H4	3.15
TS3	
Pd-N1	2.15
Pd-Cl1	2.38
Pd-Cl2	2.37
Pd-O3	2.42
C1-H4	1.15
Pd-H4	2.11
I3	
Pd-N1	2.42
Pd-Cl1	2.39
Pd-Cl2	2.42
Pd-H4	1.52
TS4	
Pd-N1	2.55
Pd-Cl1	2.81
Pd-Cl2	2.39
Pd-H4	1.56
Cl1-H4	1.71
I4	
Pd-N1	2.28
Pd-Cl2	2.38

The Pd–H4 orbital delocalizes into both σ^* antibonding Pd–Cl orbitals, and vice versa.

In the further course of the reaction, I3 undergoes reductive elimination of HCl. Our investigation shows that this step of the reaction proceeds via transition state TS4, which requires an activation energy of 78.9 kJ mol^{-1} (Table S2). The results of the IRC calculation for TS4 are presented in Fig. S6. In TS4, the Pd–H4 and Pd–Cl1 bonds are being cleaved, whereas the Cl1–H4 bond is being formed, implying that HCl leaves the reaction system. This process leads to the formation of the catalytically active Pd(0) complex, the final product of preactivation reaction (I4 in Fig. 3). NBO analysis reveals that palladium bears five lone pairs in the d orbitals, implying that the oxidative number of Pd is 0. Ligating atoms (chlorine and nitrogen) participate in bonds around palladium with more than 90%. The Pd–Cl bond delocalizes into the σ^* antibonding Pd-N1 orbital, and vice versa.

Summary

The structure of the ionic liquid [TEA][HOAc] was examined using DFT. It was shown that [TEA][HOAc] acts as multifunctional ionic liquid providing a good reaction medium, base, precatalyst-precursor, and mobile support for the active Pd species in a phosphine-free Heck reaction. The mechanism of the reaction of [TEA][HOAc] with PdCl$_2$ was examined using DFT. It was found that two

Pd(II) complexes, Pd–N,N and Pd–N,O, are formed where the Pd–N,O complex further acts as a precatalyst, yielding the catalytically active Pd(0) species. The calculated activation energies are in agreement with our experimental findings.

The reaction of PdCl$_2$ with [TEA][HOAc] provides an effective ionic liquid–palladium catalytic system for the Heck reaction. The catalytic system used is ecologically friendly and can be recovered and recycled.

References

1. Heck RF (1979) Palladium-catalyzed reactions with olefins of organic halides. Acc Chem Res 12:146–151
2. Heck FR (1985) Palladium reagents in organic synthesis. Academic, London
3. Heck FR (1991) In: Trost MB, Flemming I (eds) Comprehensive organic synthesis, vol 4. Pergamon, Oxford
4. Meijere A, Diederich F (2004) Metal-catalyzed cross-coupling reactions. Wiley-VCH, New York
5. Carmichael AJ, Earle MJ, Holbrey JD, McCormac PB, Seddon KR (1999) The Heck reaction in ionic liquids: a multiphasic catalyst system. Org Lett 1:997–1000
6. Ye C, Xiao JC, Twamley B, LaLonde AD, Norton MG, Shreeve JM (2007) Basic ionic liquids: facile solvents for carbon-carbon bond formation reactions and ready access to palladium nanoparticles. Eur J Org Chem 5095–5100
7. Wang L, Li H, Li P (2009) Task-specific ionic liquid as base, ligand and reaction medium for the palladium-catalyzed Heck reaction. Tetrahedron 65:364–368
8. Xu L, Chen W, Xiao J (2000) Heck reaction in ionic liquids and the in situ identification of N-heterocyclic carbene complexes of palladium. Organometallics 19:1123–1127
9. Pryjomska-Ray I, Trzeciak AM, Ziółkowski JJ (2006) Base-free efficient palladium catalyst of Heck reaction in molten tetrabutylammonium bromide. J Mol Catal A Chem 257:3–8
10. Henriksen ST, Norrby PO, Kaukoranta P, Andersson PG (2008) Combined experimental and theoretical study of the mechanism and enantioselectivity of palladium-catalyzed intermolecular Heck coupling. J Am Chem Soc 130:10414–10421
11. Surawatanawong P, Hall MB (2008) Theoretical study of alternative pathways for the Heck reaction through dipalladium and "ligand-free" palladium intermediates. Organometallics 27:6222–6232
12. Surawatanawong P, Fan Y, Hall MB (2008) Density functional study of the complete pathway for the Heck reaction with palladium diphosphines. J Organomet Chem 693:1552–1563
13. Petrović ZD, Marković S, Simijonović D, Petrović VP (2009) Mechanistic insight into the preactivation of a modern palladium catalyst precursor in phosphine-free Heck reactions. Monatsh Chem 140:371–374
14. Marković S, Petrović ZD, Petrović VP (2009) DFT study on the preactivation reaction of a palladium catalyst precursor in phosphine-free Heck reactions. Monatsh Chem 140:171–175
15. Goossen LJ, Koley D, Hermann H, Thiel W (2005) Mechanistic pathways for oxidative addition of aryl halides to palladium(0) complexes: a DFT study. Organometallics 24:2398–2410
16. Green JC, Herbert BJ, Lonsdale R (2005) Oxidative addition of aryl chlorides to palladium N-heterocyclic carbene complexes and their role in catalytic arylamination. J Organomet Chem 690:6054–6067
17. Cui X, Li Z, Tao ZC, Xu Y, Li J, Liu L, Guo QX (2006) N, N-dimethyl-β-alanine as an inexpensive and efficient ligand for palladium-catalyzed Heck reaction. Org Lett 8:2467–2470
18. Ahlquist M, Norrby PO (2007) Oxidative addition of aryl chlorides to monoligated palladium(0): A DFT-SCRF study. Organometallics 26:550–553
19. Li Z, Fu Y, Guo QX, Liu L (2008) Theoretical study on monoligated Pd-catalyzed cross-coupling reactions of aryl chlorides and bromides. Organometallics 27:4043–4049
20. Huang YL, Weng CM, Hong FE (2008) Density functional studies on palladium-catalyzed Suzuki-Miyaura cross-coupling reactions assisted by N- or P-chelating ligands. Chem Eur J 14:4426–4434
21. Petrović ZD, Petrović VP, Simijonović D, Marković S (2009) Mechanistic pathways for oxidative addition of aryl iodides to the low-ligated diethanolamine palladium(0) complex in phosphine-free Heck reactions. J Organomet Chem 694:3852–3858
22. Deeth RJ, Smith A, Hii KKM, Brown JM (1998) The Heck Olefination reaction; a DFT study of the elimination pathway. Tetrahed Lett 39:3229–3232
23. Graham DC, Cavell KJ, Yates BF (2006) The influence of N-substitution on the reductive elimination behaviour of hydrocarbyl-palladium-carbene complexes—a DFT study. Dalton Trans 1768–1775
24. Dupont J, de Souza RF, Suarez PAZ (2002) Ionic liquid (molten salt) phase organometallic catalysis. Chem Rev 102:3667–3692
25. Calò V, Nacci A, Monopoli A (2006) Effects of ionic liquids on Pd-catalysed carbon-carbon bond formation. Eur J Org Chem 2006:3791–3802
26. Cassol CC, Umpierre AP, Machado G, Wolke SI, Dupont J (2005) The role of Pd nanoparticles in ionic liquid in the heck reaction. J Am Chem Soc 127:3298–3299
27. Fonseca GS, Umpierre AP, Fichtner PF, Teixeira SR, Dupont J (2003) The use of imidazolium ionic liquids for the formation and stabilization of ir0 and rh0 nanoparticles: efficient catalysts for the hydrogenation of arenes. Chem Eur J 9:3263–3269
28. Hagiwara H, Sugawara Y, Isobe K, Hoshi T, Suzuki T (2004) Immobilization of Pd(OAc)$_2$ in ionic liquid on silica: application to sustainable Mizoroki-Heck reaction. Org Lett 6:2325–2328
29. Bates DE, Mayton DR, Ntai I, Davis HJ Jr (2002) CO$_2$ capture by a task-specific ionic liquid. J Am Chem Soc 124:926–927
30. Lee MK, Lee TY, Lin BJI (2003) Supramolecular liquid crystals of amide functionalized imidazolium salts. J Mater Chem 13:1079–1084
31. 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, 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, Rev A.1. Gaussian Inc, Wallingford

32. Zhao Y, Schultz NE, Truhlar DG (2006) Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. J Chem Theor Comput 2:364–382

33. Hay JP, Wadt RW (1985) Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. J Chem Phys 82:270–283

34. Petrović ZD, Simijonović D, Petrović VP, Marković S (2010) Diethanolamine and N, N-diethylethanolamine ionic liquids as precatalyst-precursors and reaction media in green Heck reaction protocol. J Mol Catal A Chem 327:45–50

35. Bouquillon S, du Moulinet d'Hardemare A, Averbuch-Pouchot MT, Hénin F, Muzart J, Durif A (1999) A coordinaton compound of (−) ephedrine and palladium(II). Acta Cryst C55:2028–2030

Comparative modeling of 25-hydroxycholesterol-7α-hydroxylase (CYP7B1): ligand binding and analysis of hereditary spastic paraplegia type 5 CYP7B1 mutations

Afraa Siam · Andrea Brancale · Claire Simons

Abstract CYP7B1 mutations have been linked directly with the neurodegenerative disease hereditary spastic paraplegia (HSP), with mutations in the CYP7B1 gene identified as being directly responsible for autosomal recessive HSP type 5A (SPG5). To evaluate the potential impact of CYP7B1 mutations identified in SPG5 on binding and protein function, a comparative model of cytochrome P450 7B1 (CYP7B1) was constructed using human CYP7A1 as a template during model construction. The secondary structure was predicted using the PSIPRED and GOR4 prediction methods, the lowest energy CYP7B1 model was generated using MOE, and then this model was assessed in terms of stereochemical quality and the side chain environment using RAMPAGE, Verify3D and ProSA. Evaluation of the active site residues of the CYP7B1 model and validation of the active site architecture were performed via molecular docking experiments: the docking of the substrates 25-hydroxycholesterol and 27-hydroxycholesterol and the inhibitor 3α-Adiol identified structurally and functionally important residues. Mutational analysis of CYP7B1 amino acid mutations related to hereditary spastic paraplegia type 5 considered phosphorylation, ligand/substrate binding and the structural roles of mutated amino acid residues, with R112, T297 and S363 mutations expected to have a direct impact on ligand binding, while mutations involving R417 would indirectly affect ligand binding as a result of impairment in catalytic function.

Introduction

CYP7B1, also called 25-hydroxycholesterol-7α-hydroxylase and oxysterol-7α-hydroxylase in reference to its physiological function, is one of two CYP7 family members. While CYP7A1 is expressed only in the liver [1–3], CYP7B1 is expressed in numerous different tissues, for example in the liver, kidney, reproductive system and in the brain, where it performs different physiological functions [4].

In the liver, CYP7B1 plays an important role in bile acid synthesis, with studies revealing the impact of this enzyme in bile acid production, particularly in early human development (rather than in adulthood). CYP7B1 also plays a significant role in atherosclerosis development, neurosteroid metabolism and sex hormone synthesis [5–7]. Many studies have revealed that the CYP7B1 enzyme may also be involved in the growth and development of the prostate and other tissues [8].

Metabolism of 27-hydroxycholesterol is carried out by CYP7B1 in the acidic pathway in the liver (Fig. 1). This metabolism is essential for achieving normal, healthy circulation with sufficient levels of cholesterol and its oxygenated derivative 27-hydroxycholesterol. Any abnormalities in CYP7B1 may lead to severe accumulation of 27-hydroxycholesterol in the circulation, thus causing hypercholesterimia and atherosclerosis. The link between

A. Siam · A. Brancale · C. Simons (✉)
Medicinal Chemistry, Welsh School of Pharmacy,
Cardiff University,
King Edward VII Avenue,
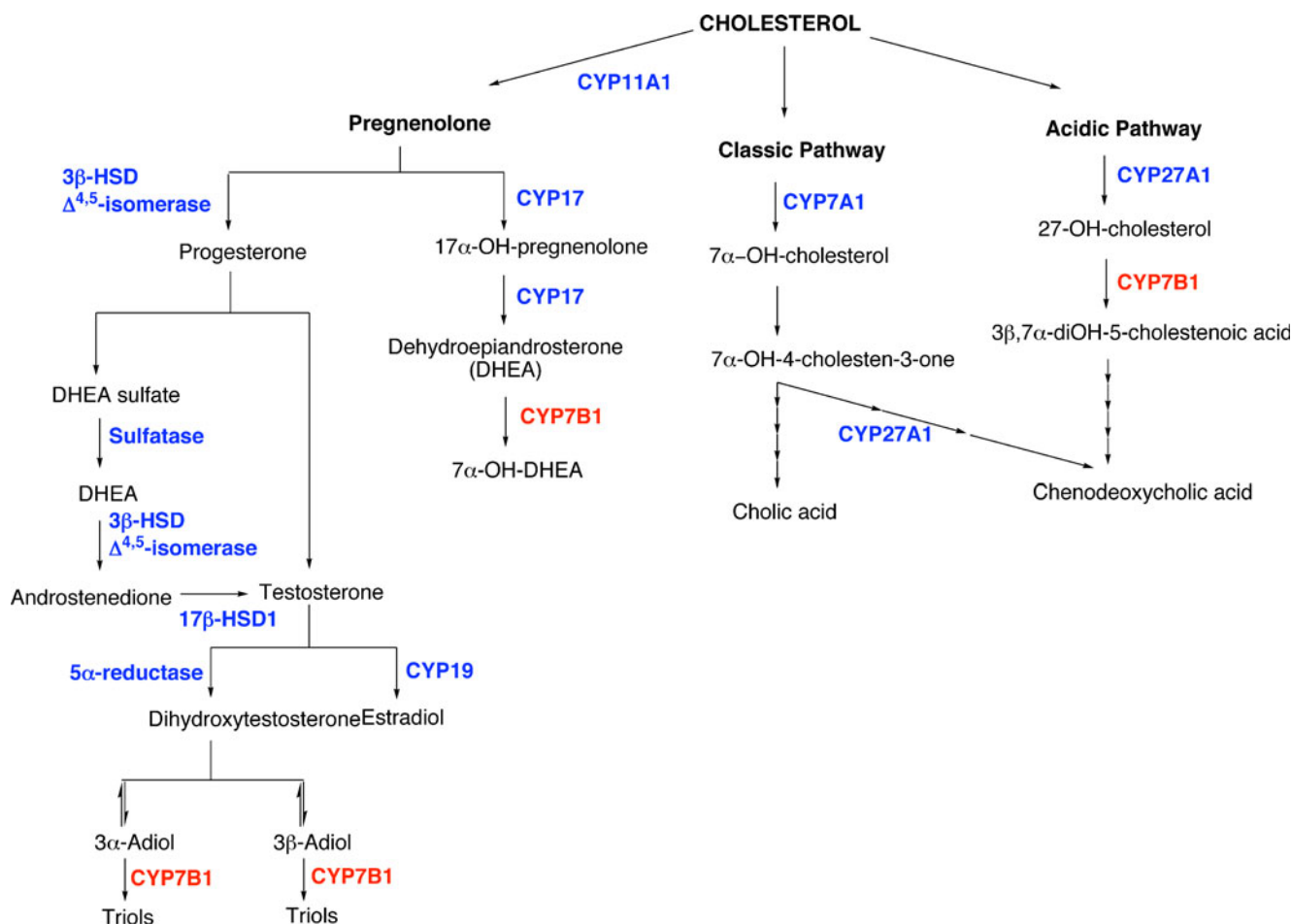Cardiff CF10 3NB, UK
e-mail: SimonsC@Cardiff.ac.uk

Fig. 1 Metabolic pathways of cholesterol

hypercholesterimia and Alzheimer's disease and the flux of 27-hydroxycholesterol in the brain is still unclear [9]. Research in this area has revealed conflicting outcomes, and reconciling them remains a great challenge. For example, Yau et al. point to low levels of CYP7B1 expression in Alzheimer's disease [10], while Attal-Khemis et al. found increased CYP7B1 formed metabolites in the sera from Alzheimer patients [11]. Indeed, Yau et al. indicate higher levels of CYP7B1 enzyme, which is the main catalyst for the 7-hydroxylation of dehydroepianderosterone (DHEA) and other sex hormone precursors in brain, play a crucial role in memory balance and cognition [10]. Studies have found that the metabolism derivative of DHEA, 7α-hydroxy-DHEA, together with other products such as 7α-hydroxy epiandrosterone (EpiA) and 7β-hydroxy-EpiA, could protect neurons from damage [12, 13], suggesting a connection between CYP7B1 hydroxylation in the brain and neuroprotection.

Studies by Pettersson et al. [7, 14] have suggested a role for the CYP7B1-mediated metabolism of 5α-androstane-3β,17β-diol (3β-Adiol) and 3α-Adiol to triols (Fig. 1) in balancing the cellular levels of hormones that activate

estrogen receptor β (ERβ) and androgen receptor (AR). As 3α-Adiol is also a powerful GABA$_A$ receptor-modulating steroid with anticonvulsant properties [15], it was speculated that CYP7B1, which is highly expressed in the hippocampus, may play a role in the control of brain levels of 3α-Adiol, and thereby influence GABA$_A$-related neuronal pathways [14].

An important recent finding linked CYP7B1 mutations directly with the neurodegenerative disease known as hereditary spastic paraplegia (HSP). Mutations in the CYP7B1 gene that are directly responsible for autosomal recessive HSP type 5A (SPG5) have been identified [16–18]. HSP is classified as a rare disease, with a recent Norwegian study carried out on a sample of 2.63 million people concluding that the number of confirmed cases of HSP was 196, corresponding to a ratio of 7.4 in every 100,000 [19]. There are three HSP diseases, which vary according to the mode of inheritance: autosomal recessive, autosomal dominance inheritance or X-linked recessive inheritance. Autosomal recessive HSP 5A or SPG5 is characterized by progressive weakness and spasticity of the lower extremities [16].

The amino acid sequence encoding CYP7B1 has been cloned from human tissue [6, 10, 20], but no crystal structures are available. Here, we describe a homology model for human CYP7B1 that provides valuable information on the active site architecture, along with docking studies that analyzed ligand-binding interactions. CYP7B1 mutations identified in SPG5 and their potential impact on binding and protein function were also evaluated.

Methods

Construction of the CYP7B1 model

Homology searching

The protein sequence of human CYP7B1 was obtained from the ExPASy server (O75881) [21]. Homologous proteins with known crystal structures were found by performing a PSI-BLAST search (comparison matrix, BLOSUM62; E-threshold, 10) using the ExPASy server, aligning the query sequence (CYP7B1) against sequences in the Protein Data Bank (PDB) [22]. The default values were used for the alignment parameters and thresholds used to screen for candidate homologues, and the BLOSUM62 comparison matrix was employed.

The Phylogeny server [23] was used to build a phylogenetic tree for the identified homologous proteins and the query sequence.

Multiple sequence and structure alignment

The secondary structure of CYP7B1 together with the templates was determined using PSIPRED [24]. The query sequence was then aligned against the most homologous templates using ClustalW [25] to identify specific α-helices, β-sheets, coils and loops, and the common features and motifs. Conserved sequences were also identified using PRATT [26].

3D model building

All molecular modeling studies were performed on an RM Intel Core 2 Duo CPU (2.2 GHz) running Windows XP and using the Molecular Operating Environment (MOE) 2008.09 [27] molecular modeling software. Homology models were built using MOE Homology using a single template approach with the AMBER99 force field [28], which uses a dictionary to set the partial charges of atoms in amino acids. Partial charges were set for the heme, which was subsequently minimized. The final homology model was constructed using human CYP7A1 (PBD 3DAX) [29] as the template. The heme coordinates were duplicated

from the CYP7A1 crystal structure and explicitly included in the homology modeling process to generate ten intermediate models. The final model was the Cartesian average of all the intermediate models. All the minimizations were performed with MOE until an RMSD gradient of 0.05 kcal mol^{-1} $Å^{-1}$ was attained with the force field specified, and the partial charges were automatically calculated.

Model validation

The stereochemical quality of the polypeptide backbone and side chains was evaluated using Ramachandran plots obtained from the RAMPAGE server [30]. The amino acid environment was evaluated using Verify3D [31], which assesses the environment of the side chain based on the solvent accessibility of the side chain and the fraction of the side chain covered by polar atoms. The ProSA tool was used to check the overall model protein structure for potential errors [32]. Validation data for the template CYP7A1 (3DAX) were used as the baseline to assess the respective models.

Docking

Two substrates and one inhibitor were used in the docking studies. The ligand structures were built using MOE Builder [27], and then the ligand was energy minimized. These ligands were docked using MOE Dock, employing Triangle Matcher as the placement method and the function London dG as the first scoring function. The refinement was set to force field and the docked poses were energy minimized in the receptor pocket. The final refined poses were ranked via MM/GBVI binding free-energy estimation (E_refine score). The receptor was set as the protein and the heme, with the binding site defined as the region within a 6.5Å radius around the heme. Thirty poses were constructed for each compound, and the best scoring model–ligand complexes were selected; the ligand interactions within these complexes were visualized using the MOE Ligand Interactions simulation.

Results and discussion

Homology searching

Comparative modeling methods use structural templates that have the highest sequence homology with the target protein. Homologous proteins were identified by scanning the protein sequence of CYP7B1, obtained from the ExPASy server [21], against 3D structures deposited in the Protein Data Bank (PDB) [16] using PSI-BLAST [33]. The search returned amino acid sequences of different

P450s. The percent sequence identities, chain lengths and E-values for the homologous protein sequences are shown in Table 1. CYP7A1 (3DAX) had the highest percent similarity (37%) and a good E-value, while the other proteins ranged from 21 to 30%. The CYP7A1, CYP8A1 and CYP3A4 sequences have similar chain lengths to the query sequence CYP7B1 (506aa), and are high-resolution structures.

To obtain more information regarding the best possible template, the Phylogeny server [23] was used to construct a phylogenetic tree using thirteen different selected proteins from CYP enzymes and other protein families. The phylogenic tree indicates that the shortest evolutionary distances occur between CYP7, CYP8 and CYP51, which confirms that CYP7, CYP8 and CYP51 constitute a single group in the P450 family (group IV), according to the evolutionary relationship and sequence similarity (Fig. 2).

Multiple sequence alignments

ClustalW [25] was used to align the suggested template sequences and the query sequence of CYP7B1 (see the "Electronic supplementary material," ESM). As all of the selected sequences were CYP subfamilies, the features common to all sequences revealed the presence of the heme binding motif, an essential part of all P450s. Moreover, by using PRATT [26], the patterns conserved across a set of sequences were identified. Twenty amino acids were nearly the same in each protein (Table 2). These 20 amino acids contained the conserved cysteine in all proteins, and this was the heme binding motif.

Alignment between the selected template (CYP7A1) and the query CYP7B1 clearly showed the cysteine ligand loop of the FXXGXXXCXG motif as a conserved region (the conserved heme-thiolate-binding signature) (Fig. 3). Using the nomenclature of Gotoh [34], six SRSs regions were identified in all selected sequence alignments based on CYP2 family analysis [35]. SRS-1 forms part of the B′-helix region, SRS-2 is within the F-helix, SRS-3 lies in a region of the G-helix, SRS-4 lies at the center of the I-helix, while SRS-5 (β1–4) and SRS-6 are just after the L-helix [36].

Secondary structure

In this study, PSIPRED [24] and GOR4 [37] (Tables 3 and 4) were used to predict secondary structure.

The ClustalW multiple alignment results for CYP7A1 and CYP7B1 show many conserved regions, which are labeled under each residue with * (see Fig. 3). The observed secondary structures of both proteins show a high degree of similarity in terms of helices and β-sheets. There is good overlap between the template's and the model's secondary structure, especially at the positions of the A-, B-, and L-helices, as well as the I-helix, which is broken in the model from residue 293 to residue 296. The model lacks the C-helix, and the template's C-helix consists of only three residues and does not form the usual conserved residue WXXR found in most P450s. There are two sets of structurally conserved β-sheets: β-sheet 1 containing five strands, and β-sheet 2 with two strands. These two sheets are considered important in the formation of the hydrophobic substrate access channel [38].

Homology model

A homology model based on the optimal template (CYP7A1) was generated using a single template alignment approach with the Molecular Operating Environment (MOE) [27] software, as described in the "Methods" section (Fig. 4). The 3D structure of the CYP7B1 model showed a high degree of similarity to the template CYP7A1, especially across the conserved motifs. The heme was sandwiched between the I-helix and the L-helix, while the conserved cysteine residue was found in the C-terminal half of the model, which acts as a fifth axial ligand for the iron of the heme cofactor.

Compared with the long I-helix in CYP7A1, the I-helix in the CYP7B1 was broken with a coil located from Ser293 to Asn296. The conserved threonine in the I-helix, thought

Table 1 The optimal crystal templates identified in the CYP7B1 BLAST results

Protein	PDB code	Blast score[a]	Sequence identity[b]	Sequence identity (%)	Chain length	E-value
CYP7A1 (human)	3DAX-A	296	178/481	37	504	1×10^{-80}
CYP8A1 (human)	2IAG-A	202	149/482	30	500	3×10^{-52}
CYP8A1 (zebra fish)	3B98-A	186	144/482	29	500	1×10^{-47}
CYP3a4 (human)	1WOE-A	69.3	68/266	25	503	3×10^{-12}
CYP51 (M. tuberculosis)	2WOA-A	65.1	53/194	27	451	5×10^{-11}
CYP120a1 (retinoic acid)	2VE3-A	46.6	96/443	21	444	2×10^{-5}

[a] The PSI-BLAST score for an alignment is calculated by summing the scores for the aligned positions and the scores for gaps

[b] The number of identical residues divided by the length of the sequence fragment identified by PSI-BLAST
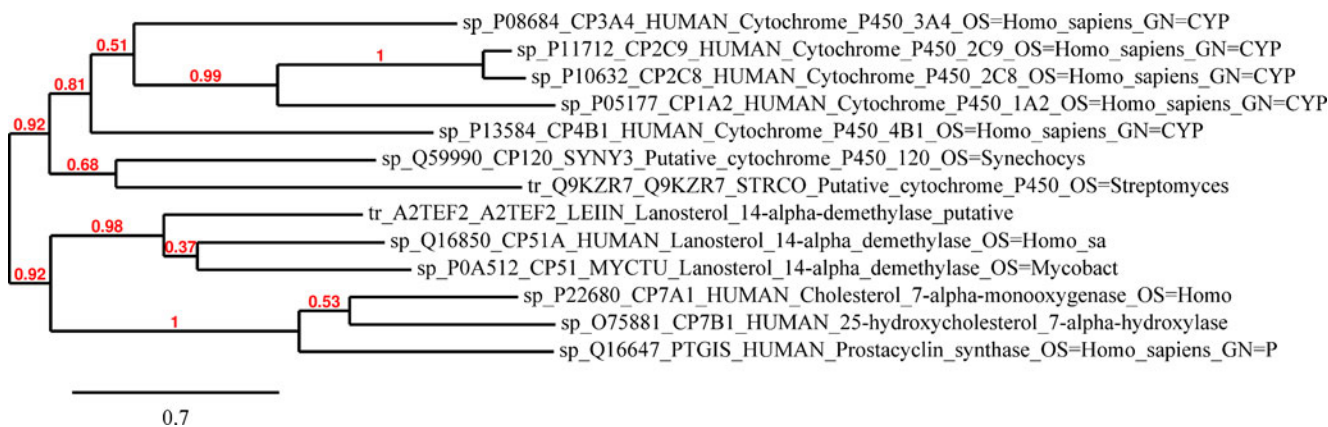
Fig. 2 Phylogenic tree generated for CYP7B1 in relation to other CYP enzymes

to be involved in the oxygen activation mechanism [39], was conserved (Thr297 in CYP7B1 and Thr290 in CYP7A1).

There are other conserved regions in P450s. These are assumed to have a significant structural function; for instance, the EXXR motif is believed to contribute to the catalytic activity of the P450 enzymes, while it also maintains the structural fold. This was found in the K-helix (Glu358 and Arg361 of the CYP7B1 3D structure). The D-helix in CYP7A1 is 15 residues longer than in the CYP7B1, while the B′ helix in CYP7B1 is five residues shorter than the B′ helix in the CYP7A1. Moreover, other α-helices (A, B, D, E and L) are very similar in the CYP7B1 model and in the CYP7A1 structure.

Interestingly, the C-helix is not found in CYP7B1, while it is composed of only three residues in CYP7A1 (Pro116, Met117 and Asp118). The C-helix in P450 enzymes is usually identified and characterized by the conserved motif WXXXR [39].

Most P450 enzymes have the important "acid–alcohol" pair located in the I-helix, with the conserved sequence (A/G)GX((E/D)(T/S). The pair is conserved with an acidic residue (Asp or Glu) just before a Thr residue, and is thought to play an essential role in the binding of an oxygen molecule for catalysis. The CYP7 family lacks this

Table 2 PRATT output showing the 20 amino acid residues conserved across all seven selected proteins

Amino acid residues	Sequence	Proteins
PFGTGtskCPGrfFAlmEIK	441–460	CYP7B1
PFGSGatiCPGrlFAihEIK	413–433	CYP7A1
PFGAGrhrCIGenFAyvQIK	384–403	CYP5A1
PWGAGhnhCLGrsYAvnSIK	411–430	CYP8A1 (human)
PFGGGlreCLGkeFArlEMK	375–394	CP120a
PFGSGprnCIGmrFAlmNMK	410–429	CP3A4
PWGTEDNLCPGRHFAVHAIK	392–412	CP8A1 (zebrafih)

conserved pair [40]. Instead, in both the CYP7A1 and CYP7B1 enzymes, the acid–alcohol pair is replaced with Ala288 and Asn289 (CYP7A1) and Ala295 and Asn296 (CYP7B1). The carbonyl oxygen of the Ala295 residue is believed to form a hydrogen bond with two water molecules, which in turn form hydrogen bonds to a residue (Leu212) located in the F-helix and to a residue in the β3–2 strand. This arrangement is very similar to that seen in the CYP8A1 (human) enzyme, in which the acid–alcohol pair is replaced with Gly286 and Asn287, and the carbonyl of Gly286 forms a hydrogen bond with Asp211 in the F-helix and a salt bridge with Leu485 in the β3–2 strand [41].

Additionally, residue Ala292 in the CYP7B1 I-helix was found to be homologous to Ala285 in CYP7A1 and Ala283 in CYP8A1 (human). This residue in CYP8A1 (human) is important in an early hydroxylation step, and this might be the case for the CYP7B1 enzyme too.

Model validation

To evaluate the quality of the modeled structures, the lowest-energy model generated from each template was subjected to a number of checks. Stereochemical quality was assessed using Ramachandran plots and the Cambridge RAMPAGE server [30], the amino acid environment was assessed using Verify3D [31] and the UCLA-DOE server [42], and the ProSA tool was used to check the overall model protein structure for potential errors [32].

Validation results suggested that the model performed well in terms of the main-chain stereochemistry and amino-acid environment. In the Ramachandran plot, 99.1% of the residues were in the allowed region, which was comparable with the corresponding value for the main template, CYP7A1 (99.8%), thus indicating that the backbone dihedral angles ψ and φ in the model were reasonably accurate. There were four outliers in the model, but they were far from the active site and do not significantly contribute to its function. Verify3D is an analytical program

```
                          Membrane        anchor         region      PolyPro
Sp|O75881|CYP7B1_HUMAN    MAGEVSAATGRFSLERLGLPGLALAAALLLLALCLLVRRTRRPGEPPLIK 50
sp|P22680|CP7A1_HUMAN     ------------------------------------SRRRQTGEPPLEN 13
                                                          *  * :.***** :

                               A'            A-helix        β1-1      β1-2      α-B
Sp|O75881|CYP7B1_HUMAN    GWLPYLGVVLNLRKDPLRFMKTLQKQHGDTFTVLLGGKYITFILDPFQYQ 100
sp|P22680|CP7A1_HUMAN     GLIPYLGCALQFGANPLEFLRANQRKHGHVFTCKLMGKYVHFITNPLSYH 63
                          * :**** .*::  :*.*::: *::**..**  * ***: ** :*:.*:

                                          β1-5    α-B`(SRS1)      α-C
Sp|O75881|CYP7B1_HUMAN    LVIKNHKQLSFRVFSNKLLEKAF---SISQLQKNHDMNDELHLCYQFLQG 147
sp|P22680|CP7A1_HUMAN     KVLCHGKYFDWKKFHFATSAKAFGHRSIDPMDGNTTEN-INDTFIKTLQG 112
                           *: : * :.:: *       ***   **. ::  *   *    : ***

                          D helixETAM EXON      helix  β3-1          E helix
Sp|O75881|CYP7B1_HUMAN    KSLDILLESMMQNLKQVFEP---QLLKTTSWDTAELYPFCSSIIFEITFT 194
sp|P22680|CP7A1_HUMAN     HALNSLTESMMENLQRIMRXXXXXXXXXXXAWVTEGMYSFCYRVMFEAGYL 162
                          ::::*: * ****:**:::.:         :* *  :*.** ::** :

                          E          F-helix   (SRS2)                  F`     α-G
Sp|O75881|CYP7B1_HUMAN    TIYGKVIVCDNNKFISELR--DDFLKFDDKFAYLVSNIPIELLGNVKSIR 242
sp|P22680|CP7A1_HUMAN     TIFGRDLTRRDTQKAHILNNLDNFKQFDKVFPALVAGLPIHMFRTAHNAR 212
                          **:*: :. :.:       *. *:* :**. *. **:.:**.:: ..:. *

                          G-helix  (SRS3)            H-helix               I helix  (SRS4)
Sp|O75881|CYP7B1_HUMAN    EKIIKCFSSEKLAKMQGWSEVFQSRQDVLEKYYVHEDLEIGAHHLGFLWA 292
sp|P22680|CP7A1_HUMAN     EKLAESLRHENLQKRESISELISLRMFLNDTLSTFDDLEKAKTHLVVLWA 262
                          **: :.:  *:* * *  :. **::. *   : :. ..:*** .  ** .***

                               I helix  (coil 293-296) in CYP7B1
Sp|O75881|CYP7B1_HUMAN    SVANTIPTMFWAMYYLLRHPEAMAAVRDEIDRLLQSTGQKKGS-GFPIHL 341
sp|P22680|CP7A1_HUMAN     SQANTIPATFWSLFQMIRNPEAMKAATEEVKRTLENAGQKVSLEGNPICL 312
                          * *****: **::: ::*:*:**** *. :*:.* *:.:*** .  * ** *

                          K helix                   (SRS5) β1-4    β2-1    β2-2
Sp|O75881|CYP7B1_HUMAN    TREQLDSLICLESSIFEALRLSSYSTTIRFVEEDLTLSSETGDYCVRKGD 391
sp|P22680|CP7A1_HUMAN     SQAELNDLPVLDSIIKESLRLSSASLNIRTAKEDFTLHLEDGSYNIRKDD 362
                          ::: :*:.* *  *.* *:****  *  .:**:** * *.* :**.*

                          β1-3  K`           the meander
Sp|O75881|CYP7B1_HUMAN    LVAIFPPVLHGDPEIFEAPEEFRYDRFIED-GKKKTTFFKRGKKLKCYLM 440
sp|P22680|CP7A1_HUMAN     IIALYPQLMHLDPEIYPDPLTFKYDRYLDENGKTKTTFYCNGLKLKYYYM 412
                          ::*::* ::* **** : *  *:***:::: **.**** .* *** * *

                          Haem pocket  L-helix           β3-3    (SRS6)
Sp|O75881|CYP7B1_HUMAN    PFGTGTSKCPGRFFALMEIKQLLVILLTYFDLEIIDDKPIG--LNYSRLL 488
sp|P22680|CP7A1_HUMAN     PFGSGATICPGRLFAIHEIKQFLILMLSYFELELIEGQAKCPPLDQSRAG 462
                          ***:*:. ****:**: **** *:::*:**:**:*:*::.     *: **

                          β4-1           β3-2
Sp|O75881|CYP7B1_HUMAN    FGIQYPDSDVLFRYKVKS-- 506
sp|P22680|CP7A1_HUMAN     LGILPPLNDIEFKYKFKHHH 482
                          :** * .*: *:**.*
```

Fig. 3 ClustalW alignment between the target sequence and the selected template structure sequence. *Asterisk* identical residues, *colon* highly conserved residues, *dot* residues are somewhat similar. All helices and β-strands are clearly labeled, and SRS regions from 1 to 6 are identified according to the Gotoh nomenclature. The *first box* shows postulated steroidogenic conserved domains, the *second box* shows the location of the active site (heme binding). The residues are colored according to their chemical properties: *red* small hydrophobic (AVFPMILWY), *blue* acidic (DE), *magenta* basic (RHK), *green* hydroxyl + amine + basic (STYHCNGQ)

that analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D) and indicates if the sequence is folded well or not. Generally, output results for 3D-1D should stay above 0.2 and should not be below 0.0 [31]. In the Verify3D output of the constructed model, 93.83% of the residues had an average

3D-1D score of >0.2, which represents a good result. The outcome for the CYP7A1 template in Verify3D indicated that 96.41% of the residues had an average 3D-1D score of >0.2.

ProSA analysis provides output in two plots. The first plot (Fig. 5a) shows the local model quality by plotting energies as a function of amino acid sequence position. In

Table 3 Comparison between CYP7B1 and the template CYP7A1

The template (CYP7A1)		The model (CYP7B1)	
Secondary structure	Residues	Secondary structure	Residues
β-sheet	Leu34–Asn36		
β-sheet	Leu38–Ileu39	β-sheet	Trp52–Leu53
A-helix	Pro52–His63	A-helix	Pro65–His76
βB1–1	Val66–Leu71	βB1–1	Phe80–Leu85
βB1–2	Lys74–Phe78	βB1–2	Lys88–Phe92
B-helix	Pro 82–Val89	B-helix	Pro96–Val103
βB1–5	Phe 95–Asp96	βB1–5	Leu109–Ser110
B′-helix	Lys99–Phe109	B′-helix	Val113–Leu118
C-helix	Pro116–Asp118	C-helix	Not found
D-helix	Ile125–Arg153	D-helix	Lys148–Lys162
Missing residues	154–164	Coil and a small helix	Gln163–Trp175
β-sheet	Val167–Gly170	β-sheet	Asp176–Glu179
E-helix	Met171–Phe188	E-helix	Leu180–Ile196
Turn	Thr193–Arg194		
F-helix	Thr195–Lys197	F-helix	Asp204–Ser228
F-helix	Ala198–Gly220		
F′-helix	Ileu225–Met227	F′-helix	Ile232–Leu234
G-helix	Arg229–Lys249	G-helix	Gly236–Leu254
H-helix	Glu255–Leu267	H-helix	Glu262–Glu272
I-helix	Asp272–Ala320	I-helix	Asp279–Ser328
Helix	321–322		(293–296, a coil)
K-helix	Gln337–Ser357	K-helix	Arg343–Leu360
β-sheet	Ser358–Ser360	β-sheet	Ser354–Ser356
β1–4	Asn362–Ala366	β1–4	Thr368–Val372
β2–1	Phe370–Leu374	β2–1	Leu376–Ser380
β2–2	Gly377–Ile381	β2–2	Gly383–Val387
β1–3	Ile386–Leu389	β1–3	Leu392–Ile395
K′-helix	Gln392–Met395	K′-helix	Pro397–His401
β-sheet	Tyr424–Cys425	β-sheet	Phe429–Lys430
β-sheet	Leu427–Leu428	β-sheet	Lys433–Lys434
β-sheet	Gly438–Ser439	β-sheet	Gly443–Thr444
β-sheet	Thr442–Ile443	β-sheet	Ser447–Lys448
L-helix	Arg447–Tyr454	L-helix	Arg452–Tyr469
β3–3	Phe465–Leu469	β3–3	Phe470–Ile474
Helix	Gln481–Arg483	Helix	Tyr484–Arg486
β4–1	Leu486–Ile488	β4–1	Phe489–Ile491
β3–2	Glu496–Phe501	β3–2	Leu499–Val504

Table 4 GOR4 output for CYP7A1 and CYP7B1

Secondary structure elements	CYP7A1	CYP7B1
Helices	204aa (40.48%)	231aa (45.65%)
β-strands	74aa (14.68%)	72 (14.23%)
Coil	226aa (44.84%)	203 (40.12%)

general, positive values correspond to problematic or erroneous parts of the input structure. The second plot (Fig. 5b) shows the overall model quality from which the z-score is calculated; its value is displayed in a plot that contains the z-scores of all experimentally determined protein chains in the current PDB. A negative score indicates a good model, whereas a positive score indicates errors. The z-score of the model was −9.56, which should be compared with the −9.16 score of CYP7A1. Further-

Fig. 4 The structure of the final CYP7B1 model with secondary structure elements (α-helices and β-strands); the meander loop is also shown in the structure located between the K ′-helix and the L-helix. The heme is shown as a *ball and stick model in gray*
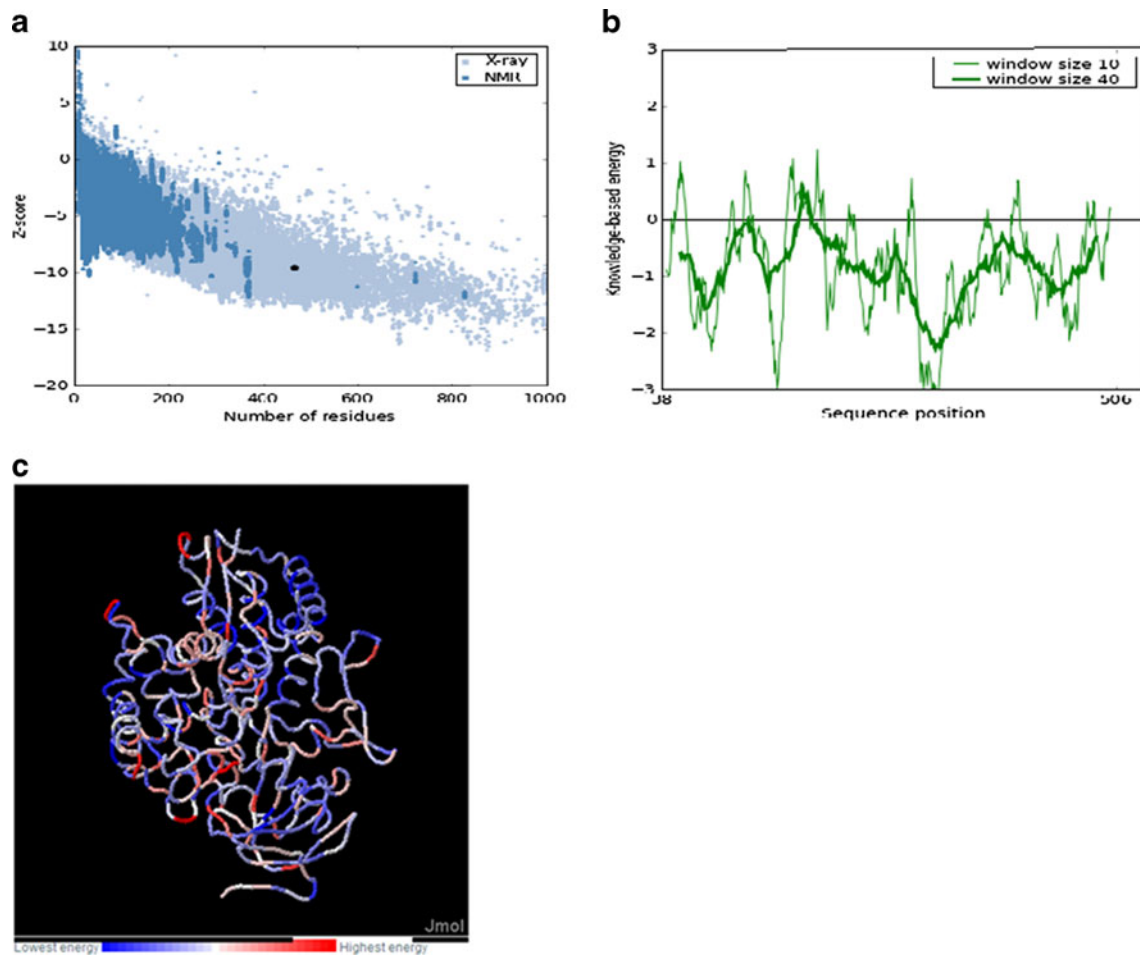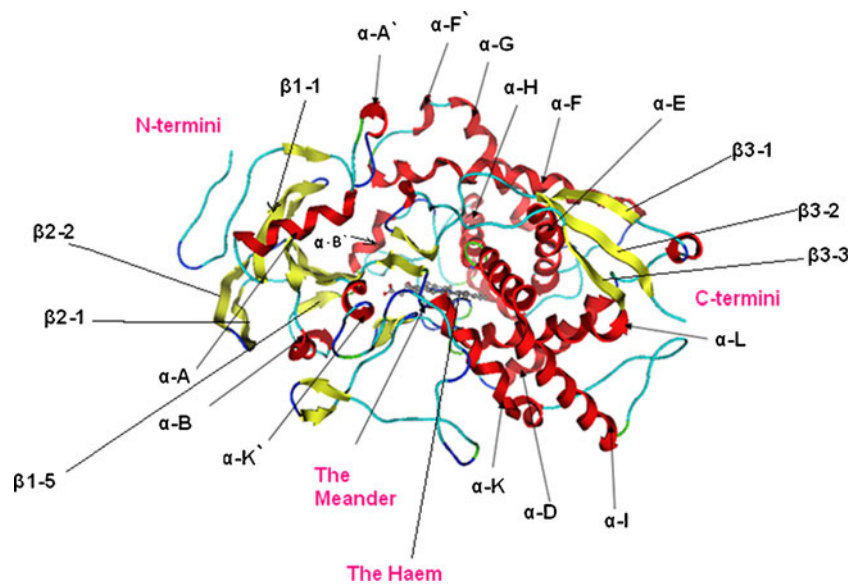
Fig. 5 a–c ProSA output for the model (CYP7B1). **a** *z*-Scores of protein chains in the PDB determined by X-ray crystallography (*light blue*) or NMR spectroscopy (*dark blue*). This plot only shows chains with <1000 residues and a *z*-score of ≤10. The *z*-score of CYP7B1 is highlighted as a *large black spot in the middle*. **b** Local model quality, shown by plotting energy as a function of amino acid sequence position; *positive values* indicate erroneous parts. Most of the amino acids gave negative values. **c** 3D structure of the CYP7B1 protein obtained using the molecule viewer Jamol

more, superimposing the model on the main template (3DAX) gave a very low RMSD (0.816Å), suggesting a high degree of similarity.

Overall, through the use of Ramachandran plots, Verify3D and ProSA analysis, a good model of CYP7B1 was generated. Further validation of the active site architecture and key binding interactions was achieved using ligand docking experiments.

Ligand docking

Although the identity between the template (CYP7A1) and the query sequence (CYP7B1) is not particularly high (37%), the functional regions of the two proteins—especially the active site—are highly conserved: approximately 66% of the heme-binding residues are conserved.

The chemical structures of the two substrates used in our docking studies were 25-hydroxycholesterol and 27-hydroxycholesterol. CYP7B1 catalyzes the 7α-hydroxylation of the 25- and 27-hydroxycholesterols formed during the first step in the alternate bile acid biosynthesis pathway [43].

25-Hydroxycholesterol was docked within the active site using MOE Dock, which generated 30 potential docking orientations. These dockings highlighted seven residues that were consistently involved in the binding of the 25-hydroxycholesterol substrate within the active site: Ser115, Leu119, Trp291, Ala292, Asn296, Ser366/Ser124 and Phe489. Arg112, Ala292, Asn296 and Ser366 were shown to form hydrogen-bonding interactions with the ligand in several docking conformations with the C7 atom positioned in proximity to the heme iron, at a distance of between 6.6 and 7.7Å. This distance would accommodate a water molecule between the 7-position of 25-hydroxycholesterol and the heme iron (Table 5).

Docking 27-hydroxycholesterol within the active site (again generating 30 potential docking orientations) highlighted eight residues that were consistently involved in the binding of the 27-hydroxycholesterol substrate within the active sites: Phe111, Arg112 and Ser115 were H-bonding residues that interacted with the C27–OH side chain; Trp291, Ala292 and Asn296 were hydrogen-bonding residues that interacted with the ring C3–OH; and Ile369, Ser447 and Phe489 formed hydrophobic interactions with the side chain. The C7 atom was positioned in proximity to the heme iron at a distance of 5.7–7.3Å (Table 5).

A recent study by the Pettersson group indicated that the CYP7B1-mediated metabolism of sex hormones, in addition to its effect on the action of estrogen, also play a role in controlling the cellular levels of androgens [14]. Furthermore, they found that 5α-androstane-3α,17β-diol (3α-Adiol) was a significant inhibitor of this enzyme, and the C-3 hydroxyl group participated in an important interaction

with the active site of the CYP7B1. 3α-Adiol was found to be able to compete with other steroids for the active site of CYP7B1 [14].

Docking 3α-Adiol generated 30 potential docking orientations, and the results indicate that the following are key residues which contribute to ligand binding: Ser115, Trp291, Ala292, Asn296, Thr367, Thr368, Ileu369, and Phe489 (Table 5). In different docking conformations, Asn296 was found to interact with C3–OH of 3α-Adiol, while Ileu369 and Thr367 were found to interact with C17–OH. C3–OH was positioned at a distance of 3.2–3.6Å from the heme iron.

Nine amino acid residues are involved in binding both the 25- and the 27-hydroxycholesterol substrates and 3α-Adiol inhibitor. All nine of these residues are located at substrate recognition sites (SRS): Ser115 at SRS-1; Trp291, Ala292 and Asn296 at SRS-4; Thr367, Thr368 and Ileu369 at SRS-5; and Leu488 and Phe489 at SRS-6. Ser115 is located in the B′-helix, and—in addition to participating in binding interactions with the ligands—may be a potential phosphorylation site (see "Mutation analysis"). Site-directed mutagenesis studies on CYP7A1 [44] showed that the large residue Trp283 (Trp291 in CYP7B1) may act as a "lid" to keep cholesterol near the heme, and that Asn288 (Asn296 in CYP7B1) was crucial to binding cholesterol in the correct orientation; it is likely that Trp291 and Asn296 in the I-helix of CYP7B1 play very similar roles. Ala292 is invariably positioned close to the C3 hydroxy group, and hydrogen-bonding interactions are expected to hold the ligand in the correct orientation above the heme (Table 5). Leu360 of CYP7A1 was implicated by site-directed mutagenesis in assisting with the orientation and positioning of cholesterol, contacting the cholesterol C18 methyl group through its terminal side-chain carbon. This interaction was preserved upon replacing leucine with isoleucine (L360I) [44]. In this region—the β1–4 sheet—of the CYP7B1 model, residues Thr367, Thr368 and Ile369 form a hydrophobic cavity in the region of the cholesterol side chain, and, in the case of 3α-Adiol, Thr367 and Ile369 undergo hydrogen-bonding interactions with C17–OH (Table 5). As previously seen with the CYP7A1 study, CYP7B1 residues Leu488 and Phe489—positioned in the β4–1 sheet—are expected to be important for correctly orienting cholesterol in the hydrophobic cavity.

Mutation analysis

In HSP5, several mutations have been identified, including nonsense mutations (R63X, R112X, Y275X, R388X) and missense mutations (G57R, F216S, T297A, S363F, R417H, R417C, F470I, R486C) (Table 6) [16, 17]. Two of the mutations have been predicted to affect the phosphorylation of CYP7B1 [16]. Using the NetPhos 2.0 server [45], it was

Table 5 Representative docking of ligands (*gray*) within the active site of CYP7B1; the key hydrogen bonding residues are shown, along with the carbon backbone in *green*
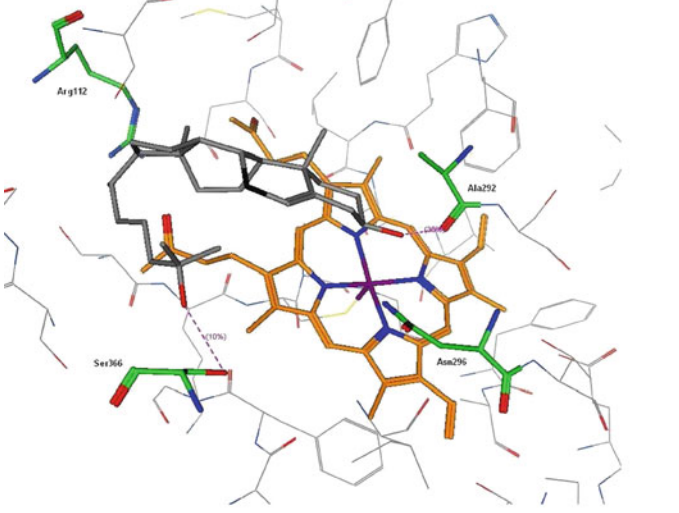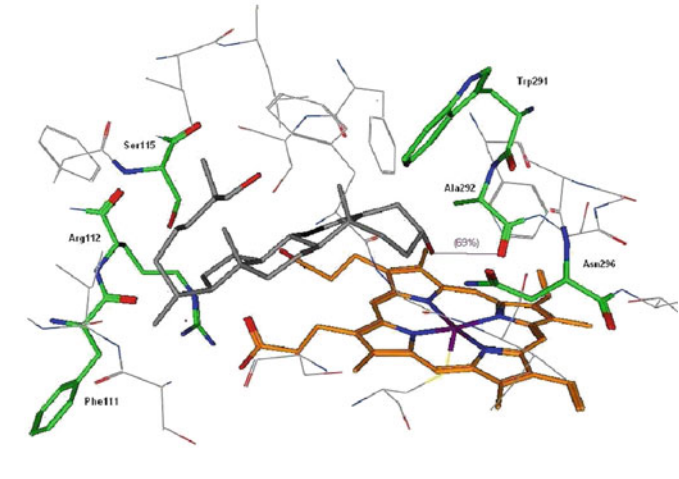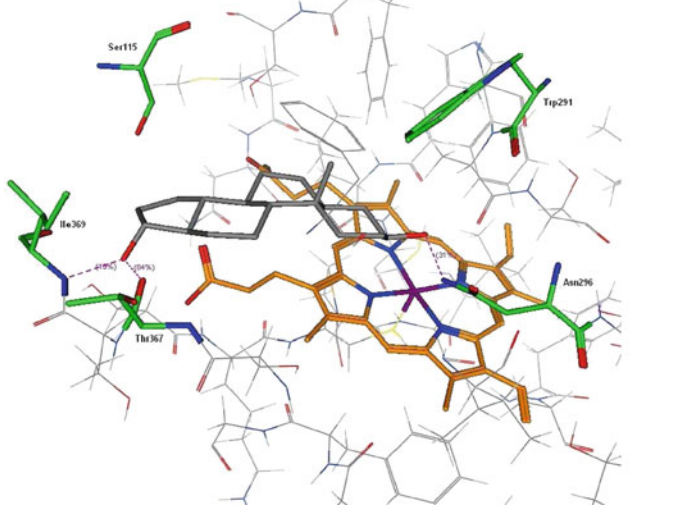
Ligand	Docking	Key interactions
25-OH-cholesterol		Arg112, Ala292, Asn296, Ser366
27-OH-cholesterol		Phe111, Arg112, Ser115, Trp291, Ala292, Asn296
3α-Adiol		Ser115, Trp291, Asn296, Thr367, Ile369

Table 6 CYP7B1 mutations identified in SPG5 families, and predicted functional effect on protein

Mutation [11, 12]	Position in protein	Functional effect on protein
G57R	Between polyproline motif and A′-helix	Structural effect on P450 tertiary fold and the A′-helix
R63X	Between A′- and A-helices	Truncation: nonfunctional protein
R112X	1aa before B′-helix, SRS-1	Truncation: nonfunctional protein
F216S	F-helix, SRS-2	Phosphorylation
Y275X	Between H- and I-helices	Truncation: nonfunctional protein
T297A	I-helix, SRS-4	Ligand binding
S363F	2aa before K-helix, SRS-5	Phosphorylation and ligand binding
R388X	Between β2–2 and β1–3	Truncation: nonfunctional protein (ligand access channel)
R417H/R417C	Meander	Interaction redox partner, structural stability
R470I	β3–3	Possible docking/access
R486C	SRS-6 helix	Ligand orientation and binding
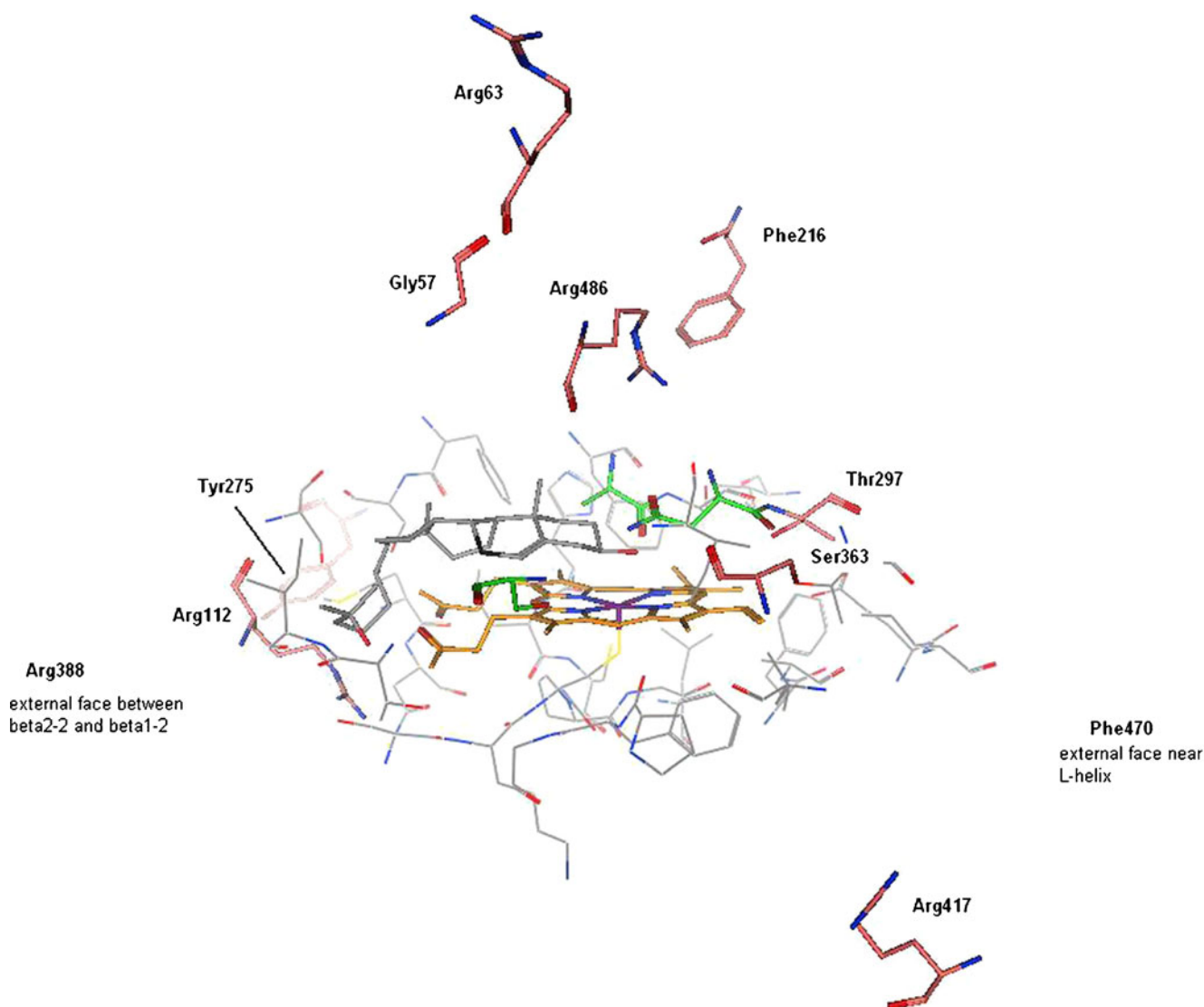
Fig. 6 Active site of CYP7B1. Mutant amino acids are identified (carbon backbone is shown in *pink*). *Orange* heme, *gray* 25-hydroxycholesterol, *green* H-bonding residues

predicted that Ser13, Ser115, Ser251, Ser363, Ser364 and Ser379 are very likely to be phosphorylation sites, with scores of >0.9. This includes residue 363 (0.985), which is mutated to a Phe in HSP (S363F). The other mutation that is affected with respect to phosphorylation is F216S. Introducing this mutation results in an additional Ser with a score of 0.997; in other words, it is very likely to be a phosphorylation site.

The mutation of G57, which is highly conserved across all P450 enzymes, is expected to have a structural effect on the protein. G57 is positioned between the polyproline motif, which has a structural role as it contributes to maintaining the P450 tertiary fold, and the A′-helix. The R112X and T297A mutations are in close proximity to key binding residues (Fig. 6), so they would impact on ligand interactions within the active site. Based on the docking studies described here, Ser115, Leu118, and Leu119 can be identified as important residues, and truncation at R112(X) would prevent the binding of the cholesterol substrates. A key binding interaction was noted between Asn296 in the I-helix and the cholesterol substrates and 3α-Adiol. Other residues in this SRS-4 region have also been implicated in ligand interactions (Trp291, Alal292, Ser293, Val294, and Ala295), so the T297A mutation would be expected to have a substantial effect on ligand interactions. Two other mutations that are likely to have an impact on ligand binding are S363F, which is in close proximity to Ile360, Ser366, Thr368, and Ile369 (identified as residues in the SRS-5 domain involved in binding interactions and the correct orientation of cholesterol), and R486C, which is in close proximity to R489 in the SRS-6 domain (Table 6, Fig. 6).

R417 resides in the meander region (the conserved PERF region) just before the cysteine pocket. The meander region has been implicated in heme binding and in interactions with redox partners for P450 reduction [35, 46, 47]. This residue may have a catalytic and/or structural function; as such, any mutation (R417H or R417C) would be predicted to severely affect enzyme function.

The four remaining mutations (R63X, Y275X, R388X, and F470I) are not obviously involved in ligand interactions based on the docking experiments, but a mutation that occurs some distance from the active site can still have a conformational effect, which may impact on the active-site architecture or the ligand access channel and subsequent binding. The nonsense mutations (R63X, Y275X, R388X) would all result in truncated sequences and a nonfunctional protein.

The R388 mutation has also been described in a disorder called CBAS3 (congenital bile acid defect type 3). This inherited mutation in the CYP7B1 gene of an infant boy was described and studied by Setchell et al. [20]. The mutation of the arginine amino acid residue R388 resulted

in premature termination (X) [6]. Truncation of the protein at this location was predicted to remove the essential heme-binding cysteine residue at 449 in the normal protein [20]. The mutation resulted in fatal hepatotoxic accumulation of 24-, 25- and 27-hydroxycholesterols and severe bleeding. The study identified that the level of 27-hydroxycholesterol alone was nearly 4,500 times more than the normal level and, as the activity level of CYP7A1 is very low in infants, it was not possible to compensate for the CYP7B1 deficiency, resulting in fatal hepatotoxicity [48].

Conclusions

The CYP7B1 model has been validated for stereochemical and amino-acid environment quality using appropriate programs, and further validation of the active-site architecture was achieved by performing docking studies with natural substrates and an inhibitor in order to determine the key residues involved in the ligand binding. In this study, many residues were found to be important residues for ligand binding: Ser115, Phe123, Ser124, Met134, Asn135, Leu140, Cys141 (in the SRS-1 region), Ile196 in the E-helix, Ala292, Ser293, Asn296, Thr297 (in the center of the I-helix, SRS-4), Ile360, Ser366, Thr368, Ile369 (SRS-5), Phe441, Ser447, Lys448, Cys449 (heme pocket), Gly451, Phe454, Ala455, Glu458, Ile459 (L-helix), and Leu488 and Phe489 (SRS6). Mutational analysis considered phosphorylation, ligand/substrate binding and the structural roles of mutated amino acid residues, and R112, T297 and S363 mutations were expected to have a direct impact on ligand binding. The positioning of R417 in the meander region leads us to predict that mutations would indirectly affect ligand binding as a result of an impairment of catalytic function due to reduced/absent heme binding and interactions with redox partners.

References

1. Myant NB, Mitropoulos KA (1977) Cholesterol 7α-hydroxylase. J Lipid Res 18:135–153

2. Noshiro M, Nishimoto M, Morohashi K, Osuda K (1989) Molecular cloning of cDNA for cholesterol 7α-hydroxylase from rat liver microsomes nucleotide sequence. FEBS Lett 257:97–100

3. Noshiro M, Okuda K (1990) Molecular cloning and sequence analysis of cDNA encoding human cholesterol 7α-hydroxylase. FEBS Lett 268:137–140

4. Stiles A, McDonald J, Bauman D, Russell D (2009) CYP7B1: one cytochrome P450, two human genetic diseases, and multiple physiological functions. J Biol Chem 284:28485–28489

5. Bjorkhem I, Andersson O, Diczfalusy U, Sevastik B, Xiu RJ, Dua C, Lund E (1994) Atherosclerosis and sterol 27-hydroxylase: evidence for a role of this enzyme in elimination of cholesterol from human macrophages. Proc Natl Acad Sci USA 91:8592–8596

6. Wu Z, Martin KO, Javitt NB, Chiang JY (1999) Structure and functions of human oxysterol 7α-hydroxylase cDNAs and gene CYP7B1. J Lipid Res 40:2195–2203

7. Pettersson H, Lundqvist J, Norlin M (2010) Effects of CYP7B1-mediated catalysis on estrogen receptor activation. Biochim Biophys Acta 1801:1090–1097

8. Tang W, Eggertse G, Chiang JY, Norlin M (2006) Estrogen-mediated regulation of CYP7B1: a possible role for controlling DHEA levels in human tissues. J Steroid Biochem Mol Biol 100:42–51

9. Heverin M, Bogdanovic N, Lutjohann D (2004) Changes in the levels of cerebral and extracerebral sterols in the brain of patients with Alzheimer's disease. J Lipid Res 45:186–193

10. Yau JL, Rasmuson S, Andrew R (2003) Dehydroepiandrosterone 7-hydroxylase CYP7B1: predominant expression I primate hippocampus and reduced expression in Alzheimer's disease. Neuroscience 121:307–314

11. Attal-Khemis SV, Dalmeyda J, Michot L, Roudier M, Morfin R (1998) Increased total 7α-hydroxy-dehydroepiandrosterone in serum of patients with Alzheimer's disease. J Gerontol 53:B125–B132

12. Li A, Bigelow JC (2010) The 7-hydroxylation of dehydroepiandrosterone in rat brain. Steroids 75:404–410

13. Chalbot S, Morfin R (2006) Dehydroepiandrosterone metabolites and their interactions in humans. Drug Metabol Drug Interact 22:1–23

14. Pettersson H, Lundqvist J, Oliw E, Norlin M (2009) CYP7B1-mediated metabolism of 5α-androstane-3α,17β-diol (3α-Adiol): a novel pathway for potential regulation of cellular levels of androgens and neurosteroids. Biochim Biophys Acta 1791:1206–1215

15. Frye CA, van Keuren KR, Erskine MS (1996) Behavioral effects of 3α-androstanediol. I. Modulation of sexual receptivity and promotion of GABA-A stimulated chloride flux. Behav Brain Res 79:109–118

16. Tsaousidou M, Ouahchi K, Warner TT (2008) Sequence alterations within CYP7B1 implicate defective cholesterol homeostasis in motor-neuron degeneration. Am J Hum Genet 82:510–515

17. Goizet C, Boukhris A, Durr A, Beetz C, Truchetto J, Tesson C, Tsaousidou M, Forlani S, Guyant-Maréchal L, Fontaine B, Guimarães J, Isidor B, Chazouillères O, Wendum D, Grid D, Chevy F, Chinnery PF, Coutinho P, Azulay JP, Feki I, Mochel F, Wolf C, Mhiri C, Crosby A, Bricel A, Stevanin G (2009) CYP7B1 mutations in pure and complex forms of hereditary spastic paraplegia type 5. Brain 132:1589–1600

18. Schule R, Suddique T, Deng HX, Yang Y, Donkervoort HM, Madrid RE, Siddique N, Schols L, Bjorkhem I (2010) Marked accumulation of 27-hydroxycholesterol in SPG5 patients with hereditary spastic paresis. J Lipid Res 51:819–823

19. Erichsen AK, Koht J, Stray-Pedersen A, Abdelnoor M, Tallaksen ME (2009) Prevalence of hereditary ataxia and spastic paraplegia in southeast Norway: a population-based study. Brain 132:1577–1588

20. Setchell KDM, Schwarz NC, O'Connell EG, Lund DL, Davis R, Lathe HR, Thompson R, Tyson W, Sokol RJ, Russell DW (1998) Identification of a new inborn error in bile acid synthesis: mutation of the oxysterol 7α-hydroxylase gene causes severe neonatal liver disease. J Clin Invest 102:1690–1703

21. Swiss Institute of Bioinformatics (2009) ExPASy (Expert Protein Analysis System) Proteomics Server. http://ca.expasy.org

22. Worldwide Protein Data Bank (2011) RCSB Protein Data Bank (PDB). http://www.rcsb.org/pdb

23. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465–W469. http://www.phylogeny.fr/

24. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292:195–202

25. 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. http://www.ebi.ac.uk/clustalw

26. Jonassen I (1997) Efficient discovery of conserved patterns using a pattern graph. Comput Appl Biosci 13:509–522

27. Chemical Computing Group Inc. (2008) Molecular Operating Environment 2008.09 (MOE). http://www.chemcomp.com

28. Weiner SJ, Kollman PA, Nguyen DT (1986) An all atom forcefield for simulations of proteins and nucleic acids. J Comput Chem 7:230–252

29. Strushkevich NV, Tempel W et al. Crystal structure of human CYP7A1, Protein Data Bank. doi:10.2210/pdb3dax/pdb

30. de Bakker P, Lovell S, Richardson Lab at Duke University (2011) RAMPAGE server. http://mordred.bioc.cam.ac.uk/~rapper/rampage.php

31. Bowie JU, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253:164–170

32. Wiederstein M, Sippl M (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410

33. 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

34. Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J Biol Chem 267:83–90

35. Lewis DFV (2001) Guide to cytochromes P450 structure and function. CRC Press, Boca Raton

36. Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J (1995) Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure 3:41–62

37. Garnier J, Gibrat JF, Robson B (1996) GOR method for predicting protein secondary structure from amino acid sequence. Meth Enzymol 266:540–554

38. Allen SW, Raucy JL (2001) Recent advances in P450 research. Pharmacogenomics J 1:178–186

39. Kim D, Heo YS, de Montellano PRO (2008) Efficient catalytic turnover of cytochrome P450(cam) is supported by a T252N mutation. Arch Biochem Biophys 474:150–156

40. Graham SE, Peterson JA (2002) Sequence alignments, variabilities and vagaries. Meth Enzymol 357:15–28

41. Chiang CW, Chan NL, Yeh HC, Wang LH (2006) Crystal structure of the human prostacyclin synthase. J Mol Biol 364:266–274

42. UCLA-DOE Institute for Genomics & Proteomics (2011) Server. http://www.doe-mbi.ucla.edu/Services

43. Pikuleva IA (2006) Cytochrome P450s and cholesterol homeostasis. Pharmacol Ther 112:761–773

44. Mast N, Graham SE, Andersson U, Bjorkhem I, Hill C, Peterson J, Pikuleva IA (2005) Cholesterol binding to cytochrome P450 7A1, a key enzyme in bile acid biosynthesis. Biochemistry 44:3259–3271

45. Blom N, Gammeltoft S, Brunak S (1999) Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol 294:1351–1362

46. Pikuleva IA, Cao C, Waterman MR (1999) An additional electrostatic interaction between adrenodoxin and P450c27 (CYP27A1) results in tighter binding than between adrenodoxin and P450scc (CYP11A1). J Biol Chem 274:2045–2052

47. Zheng YM, Baer BR, Kneller MB, Henne KR, Kunze KL, Rettie AE (2003) Covalent heme binding to CYP4B1 via Glu310 and a carbocation porphyrin intermediate. Biochemistry 42:4601–4606

48. Pullingr CR, Eng C, Salen G, Shefer S, Batta AK, Erickson SK, Verhagen A, Rivera CR, Mulvihill SJ, Malloy MJ, Kane JP (2002) Human cholesterol 7α-hydroxylase (CYP7A1) deficiency has a hypercholesterolemic phenotype. J Clin Invest 110:109–117

ORIGINAL PAPER

Geometric distortions on a three-coordinated T1 Cu site model as a potential strategy to modulate redox potential. A theoretical study

Hugo Vázquez-Lima · Patricia Guadarrama · Claudia Martínez-Anaya

Abstract A model of the three-coordinated T1 Cu site from *Trametes versicolor* was considered to evaluate the effect on redox potential of geometrical distortions in the copper coordination sphere. Systematic modifications of geometrical parameters (distances and angles) of the coordination sphere of the T1 Cu site were carried out within a density functional theory (DFT) framework, to evaluate their effects on electron affinity directly related to redox potential. The most promising result in terms of redox potential increment was distortion of the dihedral angle $C_{methylthiolate}$–S–Cu–N_{ImA} (ω), which can be rationalized as a decrease in the overlap of imidazole orbitals in the redox-active molecular orbital (β-LUMO). This overlap is minimized when ω achieves the value of 10°, therefore, this conformation might have the highest redox potential. From the molecular orbital viewpoint, a parallelism was found between the effect caused by the presence of a fourth ligand and the distorted three-coordination, which could be extrapolated to spectroscopic properties. It was also found that solvation effects on the redox potentials during geometrical distortions produce a very similar tendency, independently of the polarity of the solvent.

H. Vázquez-Lima · P. Guadarrama (✉)
Instituto de Investigaciones en Materiales,
Universidad Nacional Autónoma de México,
Apartado Postal 70–360, CU, Coyoacán,
Mexico DF 04510, Mexico
e-mail: patriciagua@iim.unam.mx

C. Martínez-Anaya
Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología (IBT), Universidad Nacional Autónoma de México, Av. Universidad 2001, Col. Chamilpa, 62250 Cuernavaca, Morelos, Mexico

Introduction

Laccases are enzymes belonging to a family known as multicopper oxidases containing four copper atoms classified in three different types (T1, T2, and T3), according to their spectroscopic properties in oxidized state [Cu (II)] [1].

These enzymes have an enormous biotechnological potential in environmental remediation since they are able to catalytically oxidize several harmful substrates such as phenols and aromatic amines. The accepted mechanism involves a first stage where Cu (II) in the T1 Cu site is reduced to Cu (I), and this active site is then re-oxidized with the concomitant reduction of O_2 to H_2O [2]. Nowadays, there is a growing interest in understanding the complex mechanisms involved in the activity of laccases and, at the same time, there is a constant search for new varieties of laccases with higher stability and higher redox potential to cover a wider spectrum of substrates [3].

The T1 Cu, responsible for the blue color of these enzymes when oxidized [4], has been studied extensively, and is a recurrent motif found in proteins involved in redox processes. This copper atom is surrounded by two histidine (His) and one cysteine (Cys) residues, all of which are conserved in all T1 Cu sites. A fourth residue can be methionine (Met), glutamine (Gln), leucine (Leu), isoleucine (Ile) or phenylalanine (Phe), depending on the protein.

In most proteins, the T1 Cu site is directly responsible for the redox potential, which is modulated by five factors: (1) the number of copper coordinated residues [lower coordination numbers of Cu (II) destabilized this state, resulting in higher redox potentials], (2) distortions in the

copper coordination sphere, (3) hydrophobicity around the copper atom, (4) electrostatic effects due to the charges of the protein, and (5) hydrogen bonds toward the sulfur atom of the coordinated cysteine [5]. Among these factors, our main interest lies in getting a deeper insight into the consequences of different distortions on the redox potential.

Most proteins containing the T1 Cu site have four residues coordinated to the copper atom. The effect of modifications in the bond distance of the fourth residue on their low/medium redox potentials has been studied widely [6]. A debate about the consequences of these structural distortions and their origins [7–9] was started in the 1990s, and some authors considered them as crucial to redox potential tuning. When the distance of the Cu-fourth residue (Met or Gln) is elongated, an increment in redox potential is observed [4]. It has been hypothesized, and theoretically supported [10–12], that the three remaining residues in T1 Cu remain almost unchanged.

On the other hand, it was observed that laccases where methionine was substituted by hydrophobic residues such as Phe or Leu, which are unable to act as fourth ligands, exhibit the highest redox potentials (regulation due to the decrease in coordination number) [3]. Thus, the combination of a low coordination number and further distortions on the three-coordinated T1-Cu sites presents itself as an attractive strategy to raise redox potentials.

Two minima, very close in energy, were found from density functional theory (DFT) calculations on simplified models of the fourth-coordinated T1 Cu site when the potential energy surface was analyzed [11], and were correlated with an amyacin mutant [10] where two different colorations can be observed; the color changes from blue to green when the protein is cooled below 200 K. Equilibrium between the two minima, assisted by small distortions of the protein, has been suggested and, depending on the enzyme, one or another state is favored. This phenomenon was rationalized in terms of the differences in the orbital interaction of the Cu–S_{Cys} bond [4]. When the interaction is dominated by a π-overlap, the protein color is blue, whereas the color turns to green when the pseudo-σ overlap controls the interaction. This classification of interactions has been described previously [13]. Most of the proteins exhibit both types of interactions, to some extent, depending on the orientation of the $d_{x^2-y^2}$ copper orbital with regard to the residues of the first coordination sphere. Particularly, in the case of *Mycelioptora thermophila* laccase, the rotation of the orbital $d_{x^2-y^2}$, when some mutations are made, promotes a change in color of the active site from blue to green, presumably due to an increase of the pseudo-σ interaction, with a concomitant decrease in the π-overlap in the Cu–S bond [14].

Taking into account a model of a three-coordinated T1 Cu site (corresponding to laccases with the highest redox potentials), and using the DFT framework, systematic modifications of geometrical parameters (distances and angles) on the coordination sphere of the active site were carried out in the present study, in order to simulate possible distortions that might result in higher redox potentials. The effect of these geometrical distortions on the electron affinity (directly related to redox potentials) was evaluated, with special emphasis on two minima characterized from the potential energy surface, linkable to an equilibrium between two states of three-coordinated copper sites.

Methods

Validation

In order to start with a good quality description of the T1 Cu site embedded in laccases, a prior validation of the theoretical methods was made on three different copper models with structural similarities to T1 Cu (see Geometric validation in the supporting information). The crystallographic structures of the chosen models available in the literature [15–17] were taken as a reference for geometrical comparisons. Among the functionals M05, BP86, B3LYP, and PBE0, available in the DFT framework of the Jaguar 7.0 program [18], the hybrid functional PBE0 [19, 20] produced the best results and was selected for further calculations (maximum deviation of 0.09 Å in the copper–ligand distances compared to crystallographic distances).

The basis set, LACVP** [21, 22] with polarized functions on all atoms was used. This is a 6-31G basis set that takes into account effective core potentials (ECPs) generated to replace the innermost core electrons for third-row (K–Cu), incorporating in this way the relativistic effects of heavy atoms.

Since the property under study is the redox potential, the chosen functional (PBE0) was evaluated on three copper complexes, mimics of the T1 Cu site (see Redox validation in the supporting information), whose redox potentials were reported in the literature [23, 24]. The PBE0 functional gives redox potential values with an average deviation of 150 mV (3.5 kcal) from experimental data. These values are within the range of the DFT redox calculation error [25].

T1 Cu model under study and methodology

One of the laccases with higher redox potentials is that from *Trametes versicolor* [3, 26]; the crystal structure of *T. versicolor* laccase deposited in the Protein Data Bank (PDB ID: 1GYC). The T1 Cu of *T. versicolor* laccase is surrounded by two histidine and one cysteine residues. In

order to systematically explore the effect of geometrical distortions (in distances and angles) on the electron affinity of the T1 Cu site, a reduced model of *T. versicolor* laccase was chosen, including two methylimidazole groups and one methylthiolate, all directly coordinated to Cu (II) (Fig. 1).

Full geometry optimization of the T1 Cu model (Cu (II), multiplicity 2) was carried out at PBE0/LACVP** level (within unrestricted formalism), and the structural deviation of the active site with regard to the complete crystallographic structure was evaluated (see Results and discussion). The full optimized T1 Cu model was used as reference for further calculations.

From the fully optimized structures of T1 Cu (Cu (II), multiplicity 2), distances A, B, and C, angles α, δ, and dihedrals θ and ω were systematically modified, fixing each one in turn and optimizing the rest of the molecule, obtaining total energies for each specific modification. To calculate the electron affinities, total energies of Cu (I) counterparts were obtained by single point calculations of each structure with the corresponding modification.

Since the geometrical restrictions proposed here do not allow a complete relaxation of the structures, the electron affinities are vertical ones. Taking into account the thermodynamic cycle of Fig. 2 [27] and the equations shown below, the redox potentials were estimated from vertical calculations of electron affinities. The reduction properties of the Cu (II) open-shell system under the unrestricted formalism are related to an unoccupied molecular orbital (labeled as β-LUMO) that was used for a qualitative description of orbitals rearrangement.
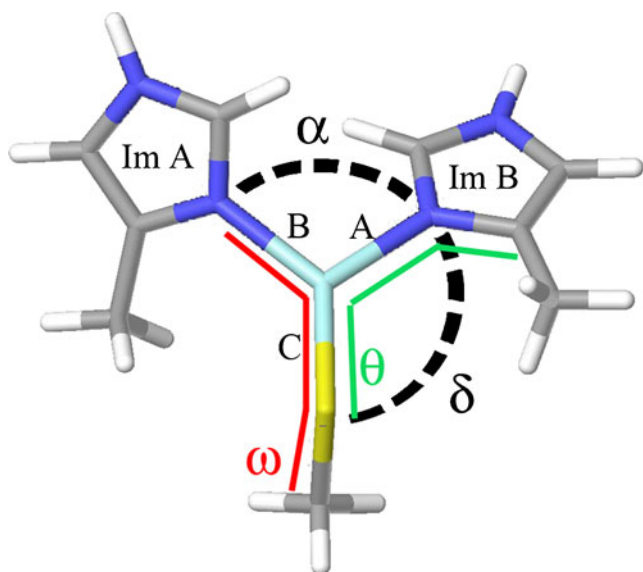
Fig. 1 T1 Cu model. Labels correspond to distances *A*, *B*, and *C*; angles α and δ, and dihedral angles ω and θ
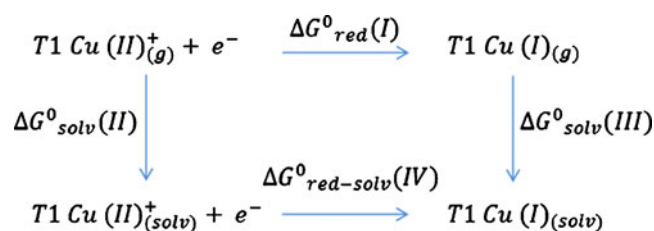
Fig. 2 Thermodynamic cycle

The global ΔG in solution ($\Delta G_{red\text{-}solv}^{0}(IV)$) can be expressed as follows:

$$\Delta G_{red-solv}^{0}(IV) = \Delta G_{red}^{0}(I) + \Delta G_{solv}^{0}(III) - \Delta G_{solv}^{0}(II) \quad (1)$$

Once the free energy change in solution is defined, the absolute redox potential E^{0} can be determined as:

$$E^{0} = -\Delta G_{red}^{0}(I)/nF - \Delta G_{solv}^{0}(III)/nF + \Delta G_{solv}^{0}(II)/nF \quad (2)$$

Where n is the number of transferred electrons and F is the Faraday constant.

The redox potential might be defined through the electron affinity in gas phase by separation of terms, $[\Delta G_{red}^{0}(I)]$ and the difference in solvation energies of the involved species $[\Delta G_{solv}^{0}(II) - \Delta G_{solv}^{0}(III)]$; thus, from the thermodynamic cycle, the electron affinity of T1 Cu (II)$_{(g)}^{+}$ is calculated as the energy of T1 Cu (II)$_{(g)}^{+}$ minus the energy of T1 Cu (I)$_{(g)}$, both in gas phase. The $\Delta G_{red}^{0}(I)$ value is estimated as the energy difference between T1 Cu (I)$_{(g)}$ and T1 Cu (II)$_{(g)}^{+}$; therefore the electron affinity of T1 Cu (II)$_{(g)}^{+}$ is equal to $-\Delta G_{red}^{0}(I)$. In this way, it is possible to establish a tendency in redox potentials through electron affinity only.

In the particular case of multicopper oxidases, there is a very small geometric reorganization of the T1 Cu site upon reduction to Cu (I) [28], according to crystallographic studies. Thus, strictly speaking, in the environment of the protein the use of ΔE values in the thermodynamic cycle would be valid (since almost no entropic contributions by reorganization are involved). Even though this approximation holds in the context of the protein and not in the small model of T1 Cu site, it can still be applied if it is considered that the geometrical distortions tested in the small model to enhance the redox potential are meant to be projected into the protein context. This approximation allows vertical electron affinities to be related to redox potentials. Therefore, ΔG values were replaced by ΔE values in the thermodynamic cycle.

Even though the most significant solvation energy is that of protein solvation (which is considered as constant), it is usually considered that the T1 Cu site is embedded in an environment with a certain dielectric constant of about 8 [29] (mostly hydrophobic); therefore, calculations in cyclohexane, tetrahydrofuran, and water were carried out

for distorted structures of the Cu (I)/Cu (II) model under the Poisson-Boltzmann solvation scheme [30, 31] and their redox potential was calculated with the aim of evaluating the possible effect of solvation on redox potential.

It has been pointed out that the solvation effects out of the coordination sphere of active sites play an important role in redox potential calculations. For most small complexes in solution, a reliable description can be obtained by means of solvation energy calculations using a continuum model that leaves aside parameters such as explicit charges, dipole moments and hydrogen bonds [32]. Some other approximations that include such parameters are: (1) QM/MM (quantum mechanics/ molecular mechanics) calculations, which explicitly take into account charges and hydrogen bonds from the protein at lower level of

theory [33]; and (2) PDLD (protein dipoles langevin dipoles), considering dipole moments and charges [34]. Beyond the method chosen, the important issue is to be aware of the limitations of every model and the conclusions that can be drawn.

Results and discussion

After full optimization of the T1 Cu model chosen in the present study, distances Cu–S, Cu–N$_{His(ImA)}$, and Cu–N$_{His(ImB)}$ (Fig. 1) are generally longer than crystallographic ones, but differing only by 0.011, 0.032, and 0.012 Å, respectively. According to these results, the protein environment apparently does not exert a significant strain on the T1 Cu site.

Fig. 3 Systematic modification of distances A (d$_A$), B (d$_B$), and C (d$_C$). d$_{ABC}$ corresponds to the synchronized modification of these three distances. The zero value in the x-axis corresponds to the distances of the fully optimized structure. a Relative energies versus distorted distances. b Vertical electron affinities versus distorted distances
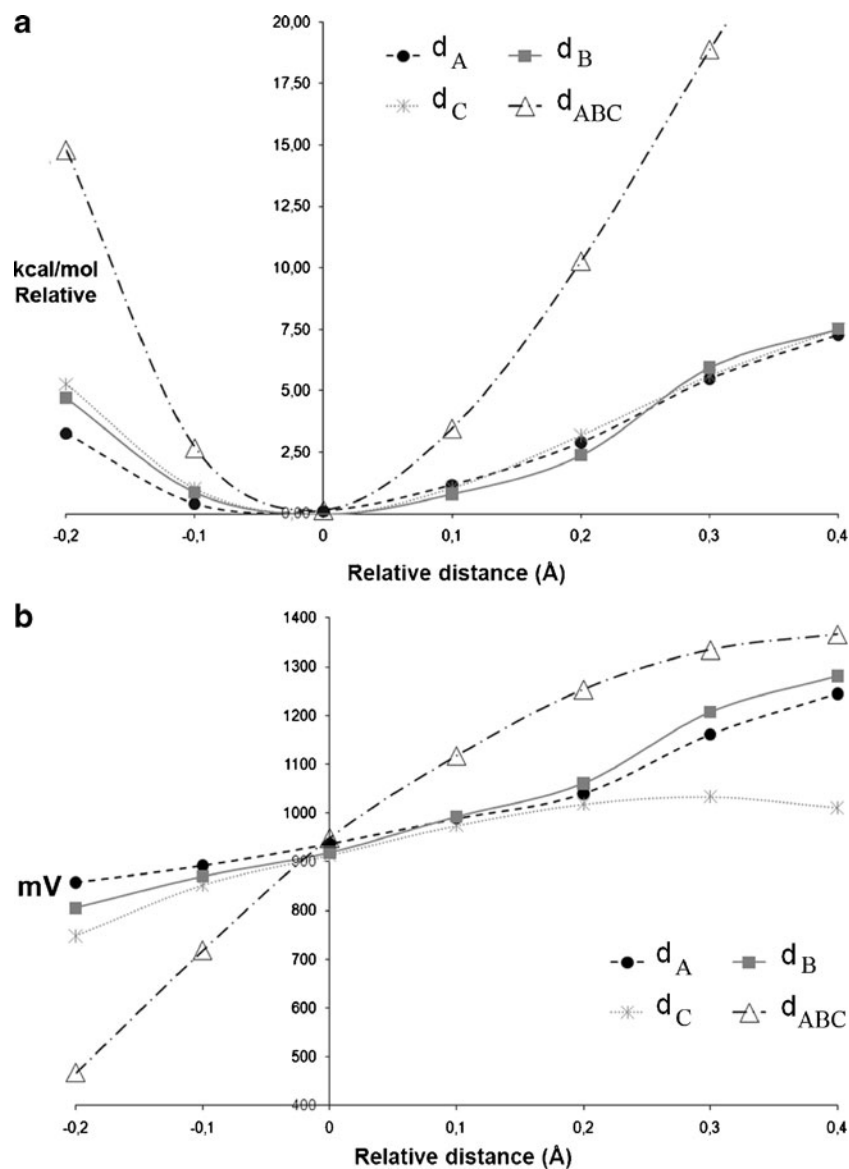
Systematic tension in bond distances

Distances A, B, and C (Fig. 3) were individually modified (shortened and increased) from −0.2 to 0.4 Å in increments of 0.1 Å. Optimization of the structures with the corresponding fixed distance were carried out and the relative energies of the complexes, as well as their vertical electron affinities (values in mV and referenced to NHE), were computed. Another series in which distances A, B, and C were synchronically modified was also generated, computing both relative energies and vertical electron affinities.

It can be seen from Fig. 3 that any tension resulting in a contraction of the distances would imply destabilization in the energy of the complex and a decrease of electron affinities.

When distances A and B (involving imidazole rings) are increased, the electron affinities also increase progressively and, even though the complex loses stability, it does so to a moderate extent. Modification of the Cu–S distance (C) resulted in a marginal increment of electron affinity, and the

synchronized movement of distances (d_{ABC}) produced the highest destabilization of the complex. However, separation of the ligands from copper, which emulates the scenario where a decrease of coordination number might occur, produces an important increment of the electron affinity. Thus, considering both effects, those tensions that increase the Cu–ligand distances might produce only a modest increment of redox potentials.

The energy of the redox active molecular orbital, labeled as β-LUMO, follows the same tendency as the vertical affinities (as shown in Fig. 4), since both parameters are related directly to the redox potentials of the T1 Cu site model; therefore, changes on this molecular orbital can be used to describe redox potential variations. This pattern was also observed in all the geometrical distortions discussed here.

Systematic distortions in angles

Next, the angles α (N–Cu–N) and δ (N–Cu–S) were independently modified from 90 to 180°, in increments of 10° and the structures were optimized with the corresponding fixed
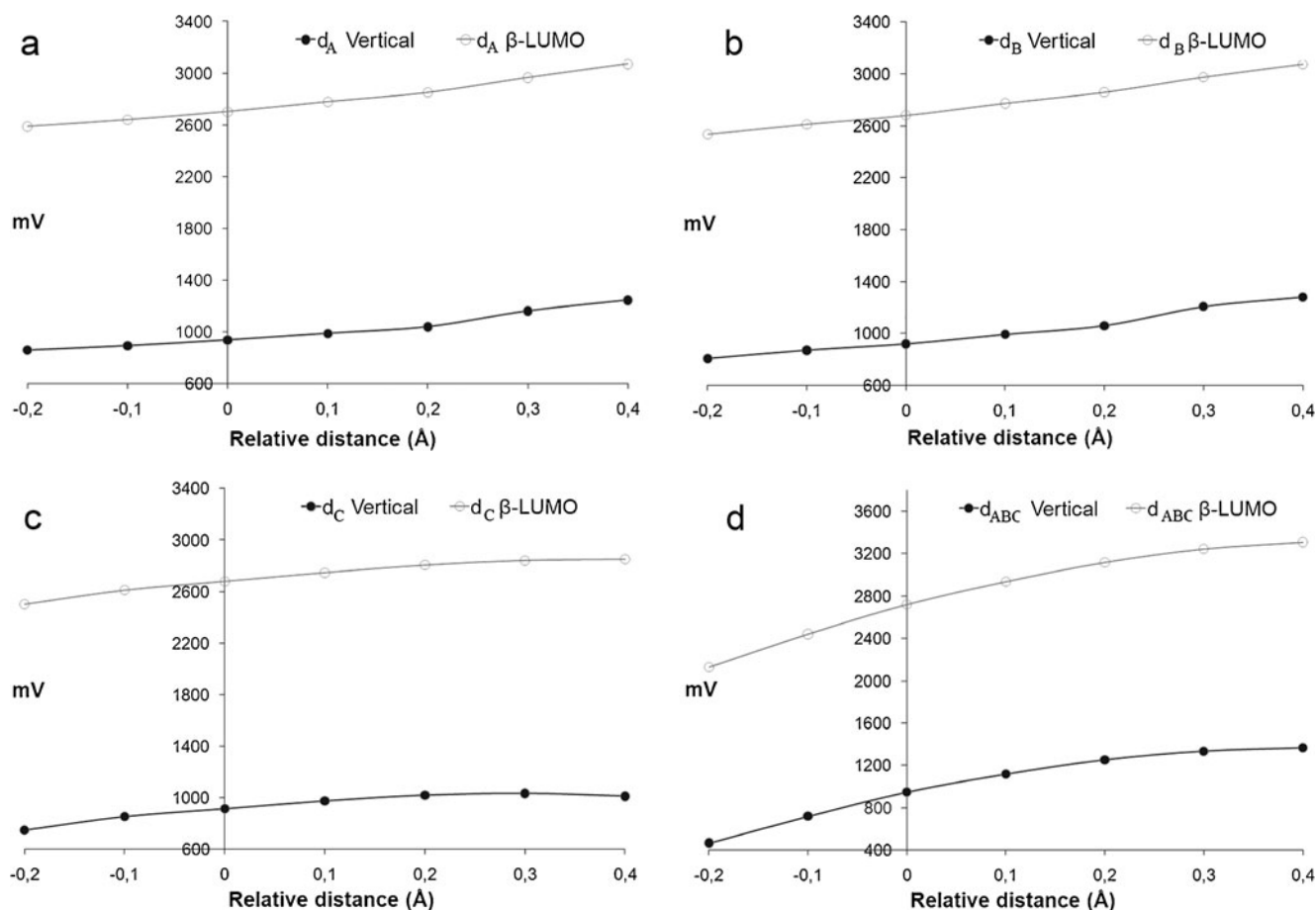
Fig. 4 Vertical electron affinities and β-LUMO energies during systematic modification of distances. **a** A, d_A; **b** B, d_B; **c** C, d_C; and **d** the synchronized distances A, B, and C. The zero value in the x-axis corresponds to the distances of the fully optimized structure
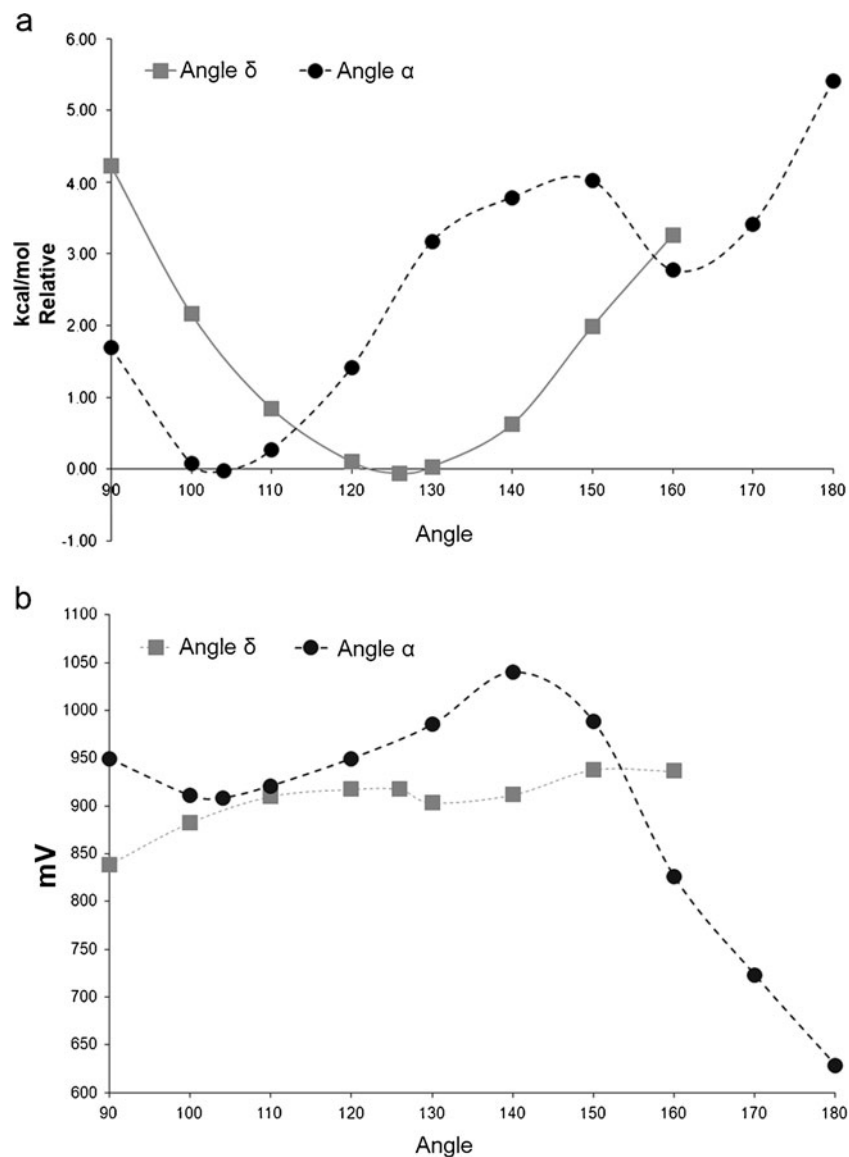
angle. The angles $\alpha=104°$ and $\delta=126°$ match with the crystallographic structure and were also considered (see Fig. 5).

As observed, the modification of angles gives rise to a very different energy pattern since additional effects, such as electronic repulsion and steric hindrance, are now involved. From the modification of angle δ, the minimum energy of the complex in the curve corresponds to the crystallographic structure. Throughout the modifications, only a small increment in the vertical electron affinity (less than 40 mV) is observed, whereas the modification of angle α produces an increment of around 120 mV, with similar complex destabilization. An increase in the angle α from 140° to 180° prevents the π interaction between copper and sulfur by steric hindrance. This effect acts in favor of the anti-bonding β-LUMO stabilization at angles of 140° and 150°, where the participation of only one of the imidazole rings via the nitrogen, and the thiolate group via a pseudo-σ

overlap is observed. This effect changes from 160° to 180° where the above scenario remains but, in addition, the overlap of the second imidazole increases, destabilizing the β-LUMO (Fig. 6).

So far a single modification of the Cu–N distance provides the best compromise between destabilization of the complex and the increment of the vertical electron affinity.

Systematic distortions in dihedral angles

The two planes involved in the dihedral angle S–Cu–N_{ImA}–C_{ImA} (θ) were defined by N_{ImA}–N_{ImB}–S–Cu atoms (first plane) and the atoms of imidazole A (second plane); see Fig. 1. Once the planes were defined, the dihedral θ was modified from 0 to 350° in increments of 10°. In this case, as imidazole A was rotating, imidazole B compensated for
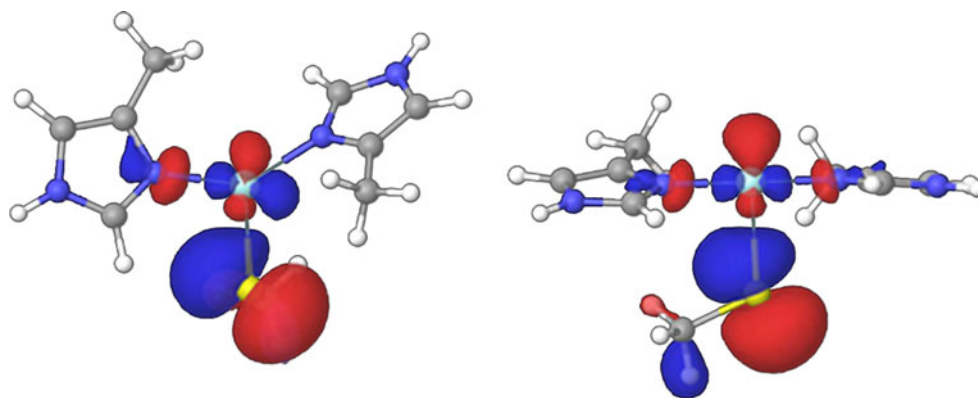
this rotation (see Fig. 7). Hence, this rotation causes a cancellation of any possible electronic rearrangement.

When the relative energies and vertical electron affinities are plotted, one can see small changes in both destabilization energy and vertical electron affinity, with a periodic behavior of both graphs (Fig. 8).

The leading role of the Cu–S interaction in the modulation of the T1 Cu redox potential has been recognized to be as a major contributor to the stabilization of the oxidized site of blue proteins [35]. To evaluate the effect of modification of the dihedral angle involving this interaction, the $C_{methylthiolate}$–S–Cu–N_{ImA} dihedral angle (ω in Fig. 1) was rotated from 0° to 270° in increments of 10°, around the plane formed by the N_{ImA}–N_{ImB}–S–Cu atoms. The relative energy of the complex and the vertical electron affinity patterns are shown in Fig. 9. It can be seen from the graph that the best trade-off between the increment of vertical electron affinity and complex destabilization is achieved with this distortion.

There is a minimum in the energy of the complex at 0° (or 180°) that is related to a change in the imidazole ring conformations. A small increment from 0° to 10° in ω

produces a change of more than 40° in the angle N_{ImA}–Cu–N_{ImB}. This result is rationalized below in terms of molecular orbitals.

The most significant increment in the vertical electron affinity is observed in conformations with ω=10° and −10°, which would correspond to a theoretical increment in redox potential of 390 mV. It is important to mention that other kinds of modification, like the presence or absence of a fourth ligand, would modify the redox potential only by 80–120 mV [4].

To explain the singular pattern of the vertical electron affinity as the dihedral angle ω is modified, the changes in the β-LUMO were represented by three key conformations selected from Fig. 9: ω=0°, 20°, and 90° (Fig. 10).

Starting with the structure where ω=90°, it is observed that the $d_{x^2-y^2}$ orbital of copper, one p orbital of sulfur, and two σ-lobes from the imidazole rings contribute to β-LUMO. When the dihedral ω is modified, the p orbital of sulfur is rotated around the Cu–S axis (see Fig. 10, lateral view, 20° configuration). The copper $d_{x^2-y^2}$ orbital follows the movement of the p orbital of sulfur to maximize the overlap between them, sacrificing the interaction between copper and the imidazole rings. The smaller overlap results in greater stability of β-LUMO and, as the dihedral ω is rotated, the β-LUMO is stabilized due to a smaller interaction with orbitals from the imidazole rings.

For ω=20°, the $d_{x^2-y^2}$ orbital from copper is out of plane by 30°, diminishing all σ interactions (Fig. 10). When the rotation is 0°, there is a notable change in the orbital disposition; the orbital $d_{x^2-y^2}$ does not contribute to β-LUMO and instead the d_{xy} orbital of copper does; this result can be interpreted also as a 45° rotation of the $d_{x^2-y^2}$ orbital around the z axis. This conformation allows the imidazole rings to interact with distal lobes of the d orbital of copper. Finally, the original p orbital of sulfur remains perpendicular to the plane of ligands due to the change in orientation, leaving the other p orbital of sulfur to interact by a single lobe with copper. This conformational change with its concomitant electronic rearrangement explains the
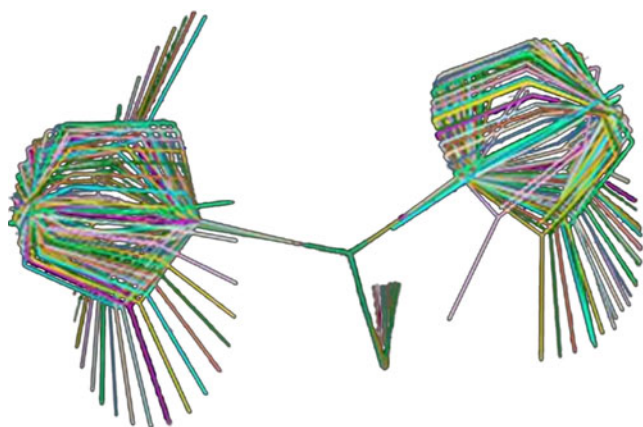
Fig. 7 Superposition of structures generated by rotation of the dihedral angle θ
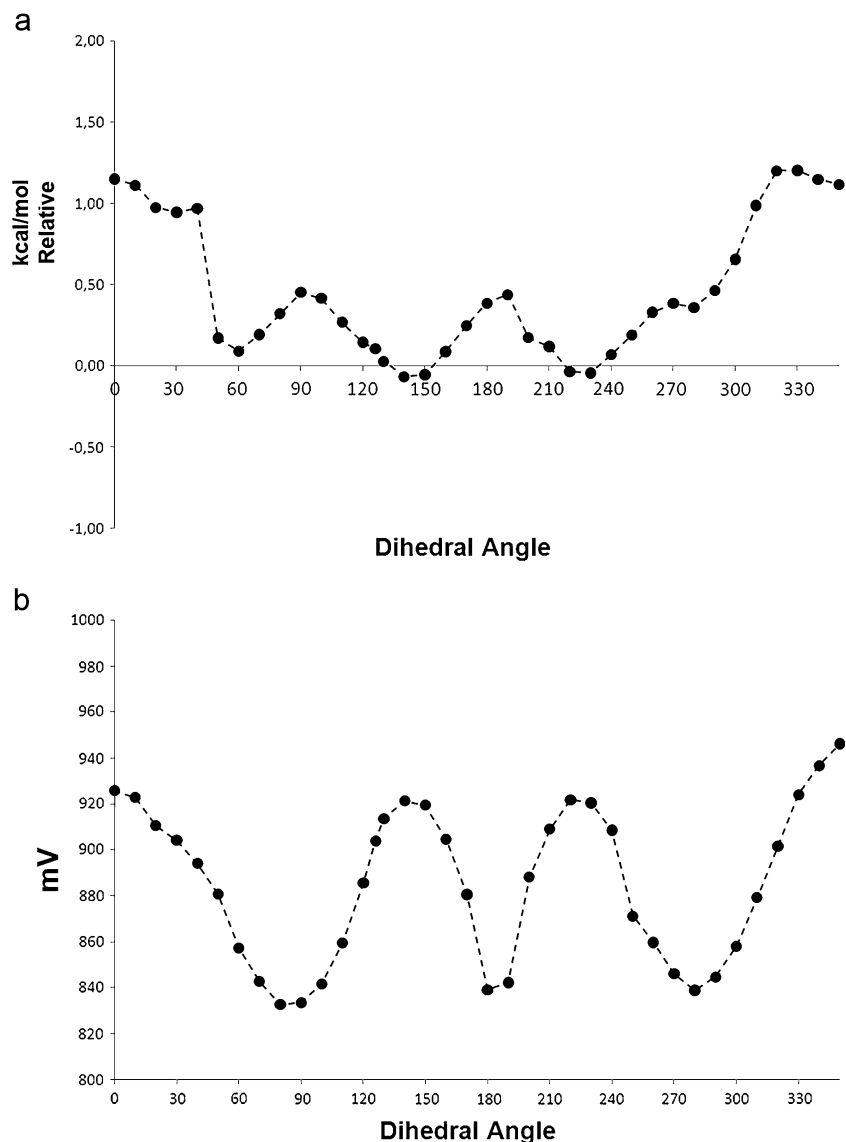
abrupt change in energy of both the complex and the β-LUMO.

Along the modification of the dihedral angle ω, two minima are generated, one in ω=90° (minimum 1) and the other in ω=0° (minimum 2). These conformations were fully optimized, and frequency calculations were carried out to verify the order of each saddle point. After optimization, the dihedral ω remains almost the same (95.3° and 0°, respectively). An energy difference of 1.7 kcal mol^{-1} between them was found, being more stable the minimum 1. The minimum 1 conformation virtually coincides with many of the crystallographic structures of high potential laccases with three-coordinated T1 Cu sites. The transition state between these minima was calculated as an energy barrier of 4.6 kcal. According to a Boltzmann distribution calculation, 6% of the T1 Cu site exists in state 2 at room temperature; however, the protein seems to

restrict the T1 Cu site to constantly adopt the conformation of minimum 1 over minimum 2. This is supported by the fact that no significant pseudo-σ interaction has been identified spectroscopically in high redox potential laccases [4], suggesting that the proteins might exert restrictions over the T1 Cu site, and that these restrictions coincide with the minimum 1.

It is important to establish that the β-LUMO in minimum 1 possesses a π-character in the Cu–S interaction, whereas in minimum 2 a pseudo-σ character prevails.

These interactions define the spectroscopic properties of the complexes. Earlier experimental studies attributed the T1 Cu site color change to an adjustment in the Cu–S interaction through the presence of a fourth ligand. According to the results obtained in the present study, the same color change can occur in the absence of a fourth ligand (see Supporting information).

Fig. 9 Systematic modification of $C_{methylthiolate}$–S–Cu–N_{ImA} dihedral angle ω. **a** Relative energies versus distorted dihedral angle, **b** vertical electron affinities versus distorted dihedral angle
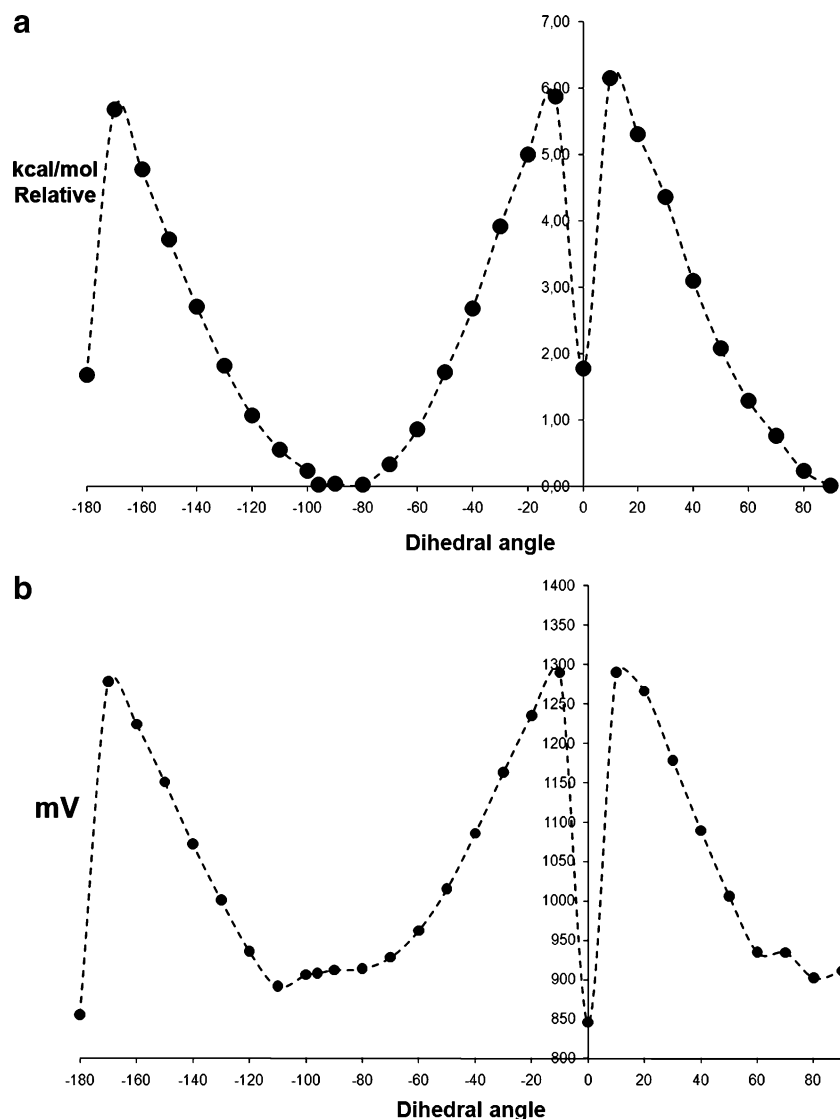
Solvation effects during systematic distortions

To evaluate the solvation effect on the redox potential of the T1 Cu site during systematic distortions, the modification of the dihedral angle ω (from 0° to 90°) was considered in three different solvents (cyclohexane, tetrahydrofuran, and water), following the thermodynamic cycle from Fig. 2. Redox potential values were estimated and are shown in Fig. 11.

It can be seen from Fig. 11 that, even though the net values of redox potential change depend on the dielectric constant of each solvent, the tendency remains almost the same. The dielectric constant inside the protein is about 8 [28], which is closer to the dielectric constant of cyclohexane (7.6). Thus, according to these results, it can be concluded that the solvation energy has little influence on any of the tendencies observed when geometric distortions are made.

Conclusions

A model of three-coordinated T1 Cu site from *T. versicolor* was used to carry out geometric distortions in a systematic fashion to evaluate their effect on redox potential modulation. Distortions in the Cu–N distance produced only a modest tuning of redox potential.

The best approach to increase redox potential was by distortion of the dihedral angle ω, rationalized as a decrease in the overlap of imidazole orbitals in the redox-active molecular orbital (β-LUMO).
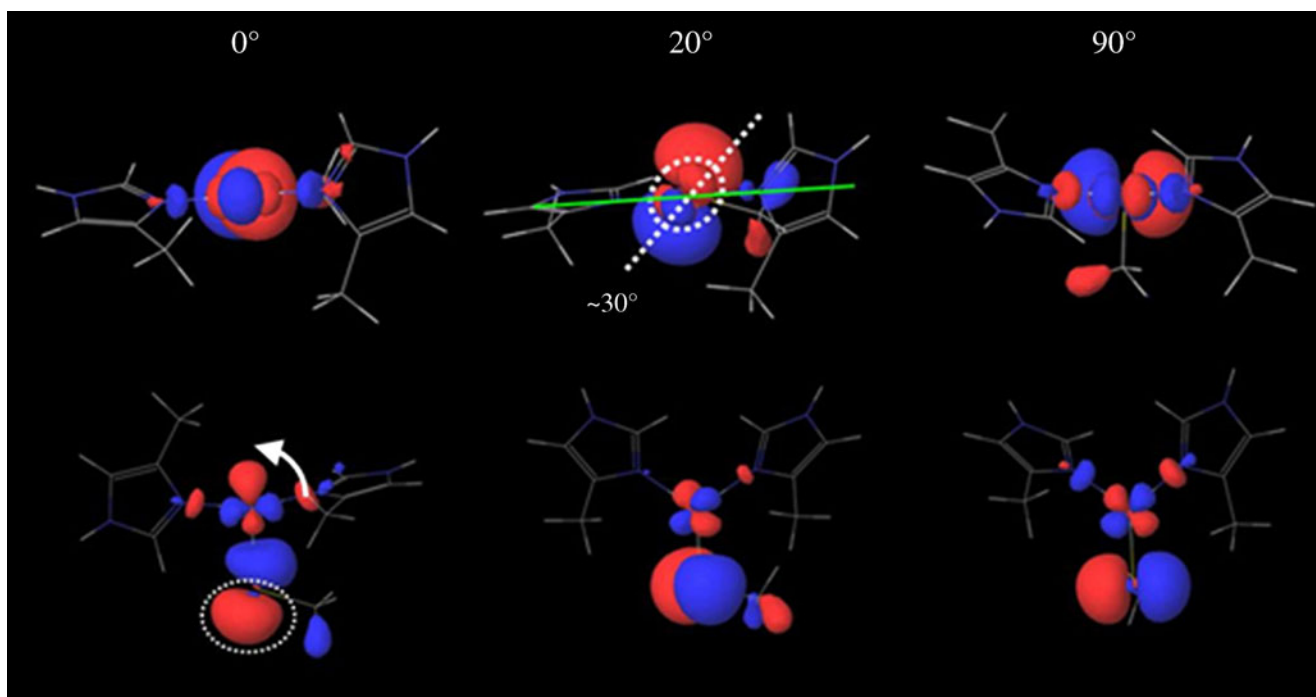
Fig. 10 β-LUMO representation in conformation with ω=0°, 20°, and 90°. *Top* Lateral view, *bottom* upper view

Another important result was the characterization of two minima, and the transition state between them for the first coordination sphere of T1 Cu site. According to the energy barrier between these two structures, they might be in equilibrium at room temperature in the absence of external restrictions (protein tensions). Hence, it is proposed that the laccase of *T. versicolor* restricts the T1 Cu site to adopt just the minimum 1 state.

The two minima described above can be distinguished clearly in terms of molecular orbitals. For minimum 1, the Cu–S interaction can be described as predominantly π in nature, whereas the minimum 2 is better described as a pseudo-σ interaction.

It was shown that the T1 Cu site redox potential during distortions follows the same pattern either in vacuum or in solution.

Fig. 11 Redox potential values of the T1 Cu model during the dihedral angle (ω) distortion in three different solvents (water, tetrahydrofuran and cyclohexane)
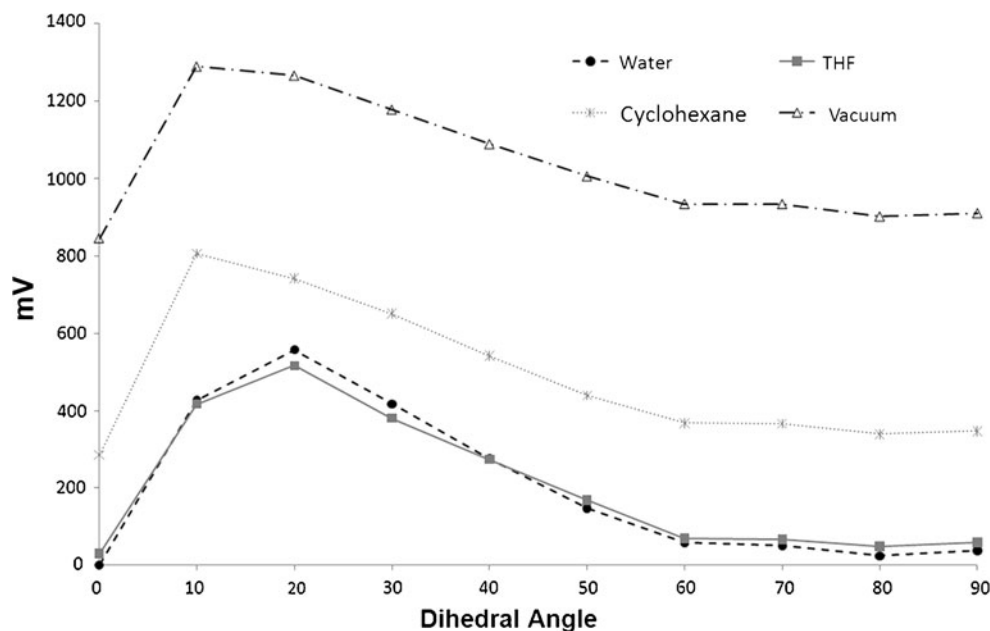
References

1. Palmer AE, Randall DW, Xu F, Solomon EI (1999) Spectroscopic studies and electronic structure description of the high potential type 1 copper site in fungal laccase: insight into the effect of the axial ligand. J Am Chem Soc 121:7138–7149. doi:10.1021/ja991087v

2. Solomon EI, Sundaram UM, Machonkin TE (1996) Multicopper oxidases and oxygenases. Chem Rev 96:2563–2605. doi:10.1021/cr950046o

3. Shleev SV, Morozova OV, Nikitina OV, Gorshina ES, Rusinova TV, Serezhenkov VA, Burbaev DS, Gazaryan IG, Yaropolov AI (2004) Comparison of physico-chemical characteristics of four laccases from different basidiomycetes. Biochimie 86:693–703. doi:10.1016/j.biochi.2004.08.005

4. Solomon EI, Szilagyi RK, George SD, Basumallick L (2004) Electronic structures of metal sites in proteins and models: contributions to function in blue copper proteins. Chem Rev 104:419–458. doi:10.1021/cr0206317

5. Li H, Webb SP, Ivanic J, Jensen JH (2004) Determinants of the relative reduction potentials of type1-copper sites in proteins. J Am Chem Soc 126:8010–8019. doi:10.1021/ja049345y

6. Christensson A, Dimcheva N, Ferapontova E, Gorton L, Ruzgas T, Stoica L, Shleev S, Haltrich YAD, Thorneley R, Aust S (2004) Direct electron transfer between ligninolytic redox enzymes and electrodes. Electroanalysis 16:1074–1092. doi:10.1002/elan.200403004

7. Ryde U, Olsson MHM, Pierloot K, Roos BO (1996) The cupric geometry of blue copper proteins is not strained. J Mol Biol 261:586–596. doi:10.1006/jmbi.1996.0484

8. Malstrom BG (1994) Rack-induced bonding in blue-copper proteins. Eur J Biochem 223:711–718. doi:10.1111/j.1432-1033.1994.tb19044.x

9. Williams RJP (1995) Energised (entatic) states of groups and of secondary structures in proteins and metalloproteins. Eur J Biochem 234:363–381. doi:10.1111/j.1432-1033.1995.363_b.x

10. Comba P, Müller V, Remenyi RJ (2004) Interpretation of the temperature-dependent color of blue copper protein mutants. Inorg Biochem 98:896–902. doi:10.1016/j.jinorgbio.2003.12.004

11. Ryde U, Olsson MHM (1999) The influence of axial ligands on the reduction potential of blue copper proteins. J Biol Inorg Chem 4:654–663. doi:10.1007/s007750050389

12. Pavelka M, Burda JV (2008) Computational study of redox active centres of blue copper proteins: a computational DFT study. Mol Phys 106:2733–2748. doi:10.1080/00268970802672684

13. Holm H, Kennepohl P, Solomon EI (1996) Structural and functional aspects of metal sites in biology. Chem Rev 96:2239–2314. doi:10.1021/cr9500390

14. Palmer AE, Szilagyi RK, Cherry JR, Jones A, Xu F, Solomon EI (2003) Spectroscopic characterization of the Leu513His variant of fungal laccase: effect of increased axial ligand interaction on the geometric and electronic structure of the type 1 Cu site. Inorg Chem 42:4006–4017. doi:10.1021/ic026099n

15. Karlin KD, Dahlstrom PL, Hyde JR, Zubieta J (1980) Structural comparison of Cu^I and Cu^{II} complexes displaying analogous N_2S_2 co-ordination; X-ray analysis of tetracoordinate [Cu(pma)]BPh$_4$

16. Brines LM, Shearer J, Fender JK, Schweitzer D, Shoner SC, Barnhart D, Kaminsky W, Lovell S, Kovacs JA (2007) Periodic trends within a series of five-coordinate thiolate-ligated [MII(SMe2N4(tren))]+(M) Mn, Fe, Co, Ni, Cu, Zn) complexes, including a rare example of a stable CuII-thiolate. Inorg Chem 46:9267–9277. doi:10.1021/ic701433p

17. Cao Y, Zheng Q, Chen C, Hu H, Huang Z (2004) Synthesis of a novel three-coordinate copper(I) complex: a structural mimic of the reduced form of type 1 site in copper protein. Inorg Chim Acta 357:316–320. doi:10.1016/S0020-1693(03)00469-9

18. Jaguar, version 7.0 (2007) Schrödinger, New York

19. Adamo C, Barone V (1999) Toward reliable density functional methods without adjustable parameters: the PBE0 model. J Chem Phys 110:6158–6170. doi:10.1063/1.3454734

20. Vetere V, Adamo C, Maldivi P (2000) Performance of the 'parameter free' PBE0 functional for the modeling of molecular properties of heavy metals. Chem Phys Lett 325:99–105. doi:10.1016/S0009-2614(00)00657-6

21. Hay PJ, Wadt WR (1985) Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. J Chem Phys 82:270–283. doi:10.1063/1.448799

22. Hay PJ, Wadt WR (1985) Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. J Chem Phys 82:299–310. doi:10.1063/1.448975

23. Holland PL, Tolman WB (1999) Three-coordinate Cu (II) complexes: structural models of trigonal-planar type 1 copper protein active sites. J Am Chem Soc 121:7270–7271. doi:10.1021/ja991533e

24. Holland PL, Tolman WB (2000) A structural model of the type 1 copper protein active site: N2S(thiolate)S(thioether) ligation in a Cu (II) complex. J Am Chem Soc 122:6331–6332. doi:10.1021/ja001328v

25. Uudsemma M, Tamm T (2003) Density-functional theory calculations of aqueous redox potentials of fourth-period transition metals. J Phys Chem A 107:9997–10003. doi:10.1021/jp0362741

26. Piontek K, Antorini M, Choinowski TJ (2002) Crystal structure of a laccase from the fungus *Trametes versicolor* at 1.90-Å resolution containing a full complement of coppers. Biol Chem 277:37663–37669. doi:10.1074/jbc.M204571200

27. Namazian M, Coote ML (2007) Accurate calculation of absolute one-electron redox potentials of some *para*-quinone derivatives in acetonitrile. J Phys Chem A 111:7227–7232. doi:10.1021/jp0725883

28. Hakulinen N, Kruus K, Koivula A, Rouvinen J (2006) A crystallographic and spectroscopic study on the effect of X-ray radiation on the crystal structure of *Melanocarpus albomyces* laccase. Biochim Biophys Acta 350:929–934. doi:10.1016/j.bbrc.2006.09.144

29. Datta SN, Sudhamsu J, Pandey A (2004) Theoretical determination of the standard reduction potential of plastocyanin in vitro. J Phys Chem B 108:8007–8016. doi:10.1021/jp0307452

30. Tannor DJ, Marten B, Murphy R, Friesner RA, Sitkoff D, Nicholls A, Ringnalda M, Goddard WA, Honig B (1994) Accurate first principles calculation of molecular charge distributions and solvation energies from ab initio quantum mechanics and continuum dielectric theory. J Am Chem Soc 116:11875–11882. doi:10.1021/ja00105a030

31. Marten B, Kim K, Cortis C, Friesner RA, Murphy RB, Ringnalda MN, Sitkoff D, Honig B (1996) New model for calculation of

solvation free energies: correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects. J Phys Chem 100:11775–11788. doi:10.1021/jp953087x

32. Evans D (2008) One-electron and two-electrontransfer in electrochemistry and homogeneous solution reactions. Chem Rev 108:2113–2144. doi:10.1021/cr068066l

33. Olsson M, Hong G, Warshel A (2003) Frozen density functional free energy simulations of redox proteins: computational studies of the reduction potential of plastocyanin and rusticyanin. J Am Chem Soc 125:5025–5039. doi:10.1021/ja0212157

34. Stephens PJ, Jollie DR, Warshel A (1996) Protein control of redox potentials of iron sulfur proteins. Chem Rev 96:2491–2513. doi:10.1021/cr950045w

35. Kitajima N (1992) Synthetic approach to the structure and function of copper proteins. Adv Inorg Chem 39:1–77. doi:10.1016/S0898-8838(08)60258-5

ORIGINAL PAPER

Combining molecular dynamics and docking simulations of the cytidine deaminase from *Mycobacterium tuberculosis* H37Rv

**Luís Fernando Saraiva Macedo Timmers · Rodrigo Gay Ducati ·
Zilpa Adriana Sánchez-Quitian · Luiz Augusto Basso · Diógenes Santiago Santos ·
Walter Filgueira de Azevedo Jr.**

Abstract Cytidine Deaminase (*CD*) is an evolutionarily conserved enzyme that participates in the pyrimidine salvage pathway recycling cytidine and deoxycytidine into uridine and deoxyuridine, respectively. Here, our goal is to apply computational techniques in the pursuit of potential inhibitors of *Mycobacterium tuberculosis CD* (*MtCDA*) enzyme activity. Molecular docking simulation was applied to find the possible hit compounds. Molecular dynamics simulations were also carried out to investigate the physically relevant motions involved in the protein-ligand recognition process, aiming at providing estimates for free energy of binding. The proposed approach was capable of identifying a potential inhibitor, which was experimentally confirmed by IC_{50} evaluation. Our findings open up the possibility to extend this protocol to different databases in order to find new potential inhibitors for promising targets based on a rational drug design process.

Keywords Free energy of binding · IC_{50} determination · Molecular docking simulation · Molecular dynamics simulation

L. F. S. M. Timmers · W. F. de Azevedo Jr.
Faculdade de Biociências, Instituto Nacional de Ciência e Tecnologia em Tuberculose (INCT-TB), Laboratório de Bioquímica Estrutural (LaBioQuest), Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS),
Av. Ipiranga 6681,
Porto Alegre, RS 90619–900, Brazil

L. F. S. M. Timmers · Z. A. Sánchez-Quitian · L. A. Basso ·
D. S. Santos (✉) · W. F. de Azevedo Jr. (✉)
Programa de Pós Graduação em Biologia Celular e Molecular,
Pontifícia Universidade Católica do Rio Grande do Sul,
Porto Alegre, RS, Brazil
e-mail: diogenes@pucrs.br

W.F. Azevedo
e-mail: walter@azevedolab.net

R. G. Ducati · Z. A. Sánchez-Quitian · L. A. Basso · D. S. Santos
Centro de Pesquisas em Biologia Molecular e Funcional (CPBMF), Instituto Nacional de Ciência e Tecnologia em Tuberculose (INCT-TB), Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS),
Av. Ipiranga 6681,
Porto Alegre, RS 90619–900, Brazil

Introduction

Mycobacterium tuberculosis (*Mt*), the causative agent of tuberculosis (TB), is one of the main causes of human death due to a single infectious agent. According to the World Health Organization (WHO) approximately two billion people (one third of the world population) are infected with *Mt*, and, based on the last WHO report, there was an estimation of 9.27 million new TB cases in 2007 [1–4]. TB is considered a neglected disease, as the majority of the cases are reported in emergent countries. However, since the 1980s, TB has returned to Europe and to the United States of America. The resurgence of TB in developed countries has been tightly related to the emergence of co-infection with the Human Immunodeficiency Virus (HIV/AIDS) [5]. Thereby, the TB/HIV synergism has become a challenge for pharmaceutical companies.

New anti-TB drugs should, ideally, present the following characteristics: lower toxicity, decrease the length of the treatment thereby improving patients' compliance, and be effective against the latent form of *Mt* allowing prevention of reactivation of the disease. These criteria have been used in structural bioinformatics approaches, mainly when the

aim is to find potential compounds that have activity against specific protein targets.

In order to find novel targets to pursuit new inhibitors against *Mt*, the pyrimidine metabolism appears to be an interesting alternative. Pyrimidine nucleotides have a pivotal role in many cellular functions, such as DNA replication and RNA transcription. These functions are regulated by the two major routes, the de novo and the salvage pathway [6]. The de novo pathway of pyrimidine biosynthesis is composed of six enzymes, responsible for the conversion of glutamine, by a series of steps, into uridine-5'-monophosphate (UMP). This set of reactions is characterized by a demand of higher amounts of energy when compared to the salvage pathway. On the other hand, the pyrimidine salvage pathway allows organisms to make use of exogenous pyrimidine bases and nucleosides that are not intermediates in the de novo pyrimidine synthesis [7–10]. One of the key enzymes in the pyrimidine salvage pathway is the Cytidine Deaminase (*CD*). *CD* is an evolutionarily conserved enzyme that participates in the recycling of cytidine and deoxycytidine into uridine and deoxyuridine, respectively (Fig. 1). The deamination reaction is coordinated by a zinc atom, which activates a water/hydroxide molecule in the initial hydrolytic attack on the C4 of the substrate. The *MtCDA* in the free form, was previously described as a homotetramer [11], as in *Bacillus subtilis*. Sang Chung and collaborators observed that, for a proper recognition of the inhibitor, the human *CD* homologue's active site is made up of residues from three or four subunits [12]. Accordingly, detailed analysis of the enzyme quaternary structure is pivotal to rational-based drug design.

In the present work, our goal was to apply computational techniques in the pursuit of potential inhibitors of *MtCDA* enzyme activity. Molecular docking simulation was employed to both find hit compounds and to rank the best fit of the ligands. Molecular dynamics (MD) simulations were performed in two steps: First, to verify the stability of the quaternary structure previously published [11], comparing it with the structure generated by the symmetry operators. Second, extract physically relevant motions that are essential and presumably meaningful for the protein-ligands recognition processes in the binding pocket, using *principal component analysis* (PCA). We have also calculated the free energy of binding to correlate in silico and in vitro results.

Materials and methods

Validation of the molecular docking simulation approach

Molecular docking simulation was performed to find potential inhibitors of *MtCDA* enzyme activity. Prior to this, however, the search algorithm and the scoring function, implemented in the AutoDock 4.0.2, were tested in order to confirm that this program is capable of ranking the best potential inhibitors. Initially, a redocking was carried out with the structure 3LQP (PDB access code) aiming at verifying whether the program could reproduce the ligand location found in the crystallographic structure. Then, we tested the ability of the program to identify, qualitatively, the best inhibitors previously described [12–14]. Considering that there is no inhibitor described for *MtCDA*, we used the human *CD* (PDB access code: 1MQ0) [12]. Table 1 presents the result for the qualitative validation of the human *CD* based on the free energy of binding calculated with the AutoDock 4.0.2 and the constants determined in vitro.

Database construction

The database consists of 91 compounds, where 84 molecules are analogs of the substrates, cytidine and
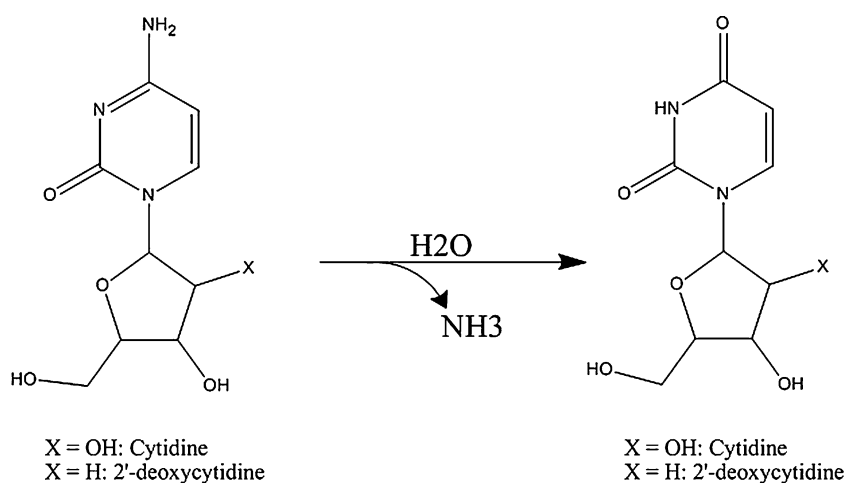
Fig. 1 *MtCDA* chemical reaction

X = OH: Cytidine
X = H: 2'-deoxycytidine

X = OH: Cytidine
X = H: 2'-deoxycytidine

Table 1 The result for the qualitative validation of the human *CDA* based on the free energy of binding calculated with the AutoDock 4.0.2 and the constants determined *in vitro*

Inhibitor	$K_{i(mM)}$	*ΔG_{exp}	ΔG_{calc}
Diazepinone	2.5×10^{-8}	-10.78	-6.41
Tetrahyuridine	9.5×10^{-7}	-8.54	-5.81
Zebularine	6.7×10^{-6}	-7.34	-5.79
CMP	0.39×10^{-3}	-4.83	-5.15

*$\Delta G = R \times T \times \ln K_i$

deoxycytidine. Among these analogs, 49 were obtained from Pubchem, 25 from DrugBank [15], and 10 from ZINC [16] databases. Tanimoto coefficient cuttof of 90% was used to search the analogs. The other seven compounds were constructed based on modifications of the substrates and products, where the amine and hydroxyl groups were substituted by methyl, and sulfur, and the pentose ring was modified for an aliphatic chain.

Molecular docking simulation

The flexible docking simulations were performed using PyrX-0.5, where the AutoDock 4.0.2 is implemented. The AutoDock software uses an empirical scoring function based on the free energy of binding. Among the stochastic search algorithms offered by the AutoDock suite, we chose the Lamarckian genetic algorithm (LGA) that is a hybrid approach [17], which combines genetic algorithm (as global search) [18] and Solis and West algorithm (as local search) [19].

The *MtCDA* structure used for the simulations was 3LQP (PDB access code). A grid box was created with $36 \times 31 \times 38$ points and a resolution of 0.375 Å in order to include solely the protein's active site, aiming to reduce the computational cost. The molecular docking process was carried out with 100 independent runs for each docking simulation, an initial population of 150, a maximum number of 500,000 energy evaluation, and a maximum number of 27,000 generations. Mutation and crossover were applied to the population at rates 0.02 and 0.80, respectively.

Molecular dynamics simulations

All MD simulations were carried out using the GROMACS 4.0.5 package [20] with the 53a6 GROMOS force field. We performed MD simulations for the enzyme in the free form (PDB access code: 3IJF), with the products (uridine and deoxyuridine - PDB access code: 3LQP and 3LQT, respectively), and the ten best results obtained from the molecular docking simulations, in order to investigate the dynamic nature of the interactions between the protein-ligand complexes, and water molecules.

The MD simulations were carried out using the particle mesh Ewald method [21] for the electrostatic interactions. The van der Walls and Coulomb cutoff were 14 Å and 10 Å, respectively. The integration time step was 2.0 fs, with the neighbor list being updated every ten steps by using the grid option and a cutoff distance of 12 Å. The simple point charge extended (SPC/E) [22] water model was used. Periodic boundary condition has been used with constant number of particles, pressure, and temperature (NPT) in the system. The V-rescale thermostat was applied using a coupling time of 0.1 ps to maintain the systems at a constant temperature of 298.15 K. The Berendsen barostat was used to maintain the systems at a pressure of 1 bar, and values of the isothermal compressibility were set to 4.5×10^{-5} bar^{-1} for water simulations. The temperature of the systems were increased from 50 to 300 K in five steps (50–100, 100–150, 150–200, 200–250, and 250–300 K), and the velocities at each step were reassigned according to the Maxwell-Boltzmann distribution at that temperature and equilibrated for 10 ps, except for the last part of the thermalization phase, which was of 40 ps. The systems were submitted to a steepest descent followed by conjugated gradient energy minimizations up to a tolerance of 1000 kJ mol^{-1}. An MD simulation with position restraints was carried for a period of 20 ps in order to allow the accommodation of the water molecules in the system. Finally, 10 ns MD simulations were performed to all systems. The topologies files and other force field parameters, except the charges of ligands, were generated using the PRODRG program [23]. The partial atomic charges to the ligands were calculated using the Gaussian03 package [24], being submitted to single-point ab initio calculations at DFT/B3LYP/6-311 G (2d, p) level in order to obtain ESP charges.

Principal component analysis

Essential dynamics (ED), also known as PCA, is an interesting method for the identification of the main conformational changes, which often have importance in biological process for a protein during MD simulation. The ED analysis is a technique that reduces the complexity of the data and extracts the concerted motion in simulations that are essentially correlated and presumably meaningful for biological function [25]. In the ED analysis, a variance/covariance matrix was constructed from the trajectories after removal of the rotational and translational movements. A set of eigenvectors and eigenvalues was identified by diagonalizing the matrix. The eigenvalues represented the amplitude of the eigenvectors along the multidimensional space, and the displacement of atoms along each eigenvector showed the concerted motions of protein along each direction. An

Table 2 The free energy of binding calculated in silico compared to experimentally determined

Inhibitors	Lennard-Jones^{vdW}	Coulomb^{el}	ΔG_{bind}	ΔG_{bind} Exp
Zebularine	-22,972	-94,240	-8,954	-8,541
Dihydrouridine	-35,051	-23,660	-6,377	-6,237
Fluorozebularine	-30,946	-25,720	-5,923	-9,251
Diazepinone	-60,552	-54,550	-11,846	-10,782
Tetrahydrouridine	-40,392	-9,130	-6,249	-7,338

*The ΔG_{bind} values are represented in *Kcal*

assumption of ED analysis is that the correlated motions for the function of the protein are described by eigenvectors with large eigenvalues. The movements of protein in the essential subspace were identified by projecting the Cartesian trajectory coordinates along the most important eigenvectors from the analysis.

Free energy of binding calculation and determination of the α and β coefficients

The linear interaction energy approximation (LIE) method [26] was employed to calculate the free energy of binding to all protein-ligand complexes. The LIE method is a semiempirical approach that combines advantages of free energy perturbation (FEP) and thermodynamics integration (TI) in order to obtain the free energy of binding. Often, two MD simulations are required to calculate the free energy of binding with the LIE method: one for the free ligand in solution and another for the protein-ligand in solution. The LIE equation is as follows:

$$\Delta G_{bind} \approx \alpha \Delta \langle V_{l-s}^{vdW} \rangle + \beta \Delta \langle V_{l-s}^{el} \rangle + \gamma, \quad (1)$$

where $\langle \rangle$ denotes MD averages of the non-bonded van der Walls (*vdW*) and electrostatic (*el*) interactions between the ligand and its surrounding environment (*l-s*). The α and β values are coefficients that represent nonpolar and polar contributions, respectively, and γ is a constant term.

We performed four simulations for each protein-ligand complex: two simulations with the partial atomic charges of the ligand, as described above, and another two simulations without charges to the ligands in order to obtain the long-range electrostatic values. Before initiating the LIE calculations with the ten best docking results, we calibrated the coefficients based on the structure of the human *CDA* (PDB access code: 1MQ0) [12], which has inhibitory constants experimentally calculated for five different inhibitors. The coefficients applied to the LIE calculations were α=0.59, β=0.43, and γ=0, which obtained a correlation coefficient of 0.73 between experimental and calculated values, and the average to the estimated error was 13.29% or $\pm 3\,kcal \times mol^{-1}$ (Table 2), similar to that considered in the AutoDock [17].

IC$_{50}$ determination

In vitro studies with the human *CDA* have shown that tetrahydrouridine inhibits the enzyme-catalyzed chemical reaction with a K_i value of 6.7 μM [27]. In the present study, attempts have been made to investigate the level of tetrahydrouridine inhibition of *MtCDA* enzyme activity in vitro. The evaluation of IC$_{50}$ value, concentration of inhibitor required to effect 50 % reduction en enzyme activity, is a good approach to assess relative inhibitor potency. Thereby, tetrahydrouridine IC$_{50}$ for *MtCDA* was determined by measuring initial rates at fixed non-saturating levels of cytidine (K_M=1004 μM [11]) in either absence or presence of tetrahydrouridine (0–1400 μM) in the reaction mixture, being the maximal reaction rate condition determined in the absence of inhibitor. It should be pointed that each individual initial rate datum was the average of duplicate or triplicate measurements. The IC$_{50}$ value, obtained by fitting the data to the appropriate equation (Eq. 2) using SigmaPlot 2004 (Systat Software, Inc.), defines the concentration of inhibitor required to half-saturate the enzyme population; v_i and v_o are, respectively, the reaction velocity in the presence and absence of inhibitor, and v_i/v_o represents the fractional activity remaining at a given inhibitor concentration (fraction of free enzyme) [28].

$$\frac{v_i}{v_o} = \frac{1}{1 + \left(\frac{[I]}{IC_{50}} \right)} \quad (2)$$

Table 3 The minimum, mean, maximum, and standard deviation of the RMSD and RG between structures A and B

Analysis		Structure A	Structure B
RMSD	Min	0.001202	0.001248
	Max	0.501063	1.18391
	Mean	0.380708	0.605804
	Stdev	0.0680289	0.236939
RG	Min	2.12377	2.48344
	Max	2.3302	3.15195
	Mean	2.29459	2.80797
	Stdev	0.024201	0.150226

*The values are represented in "*nm*"

Fig. 2 Graphical representation of RMSD and RG of each *MtCDA* tetrameric structure (**a** and **b**) as a function of time
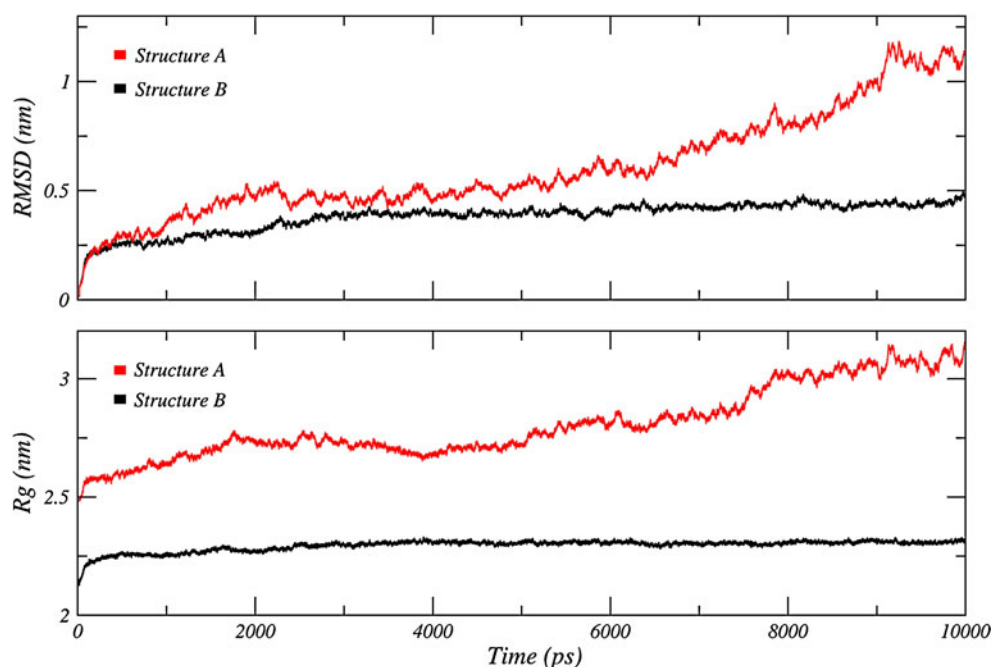
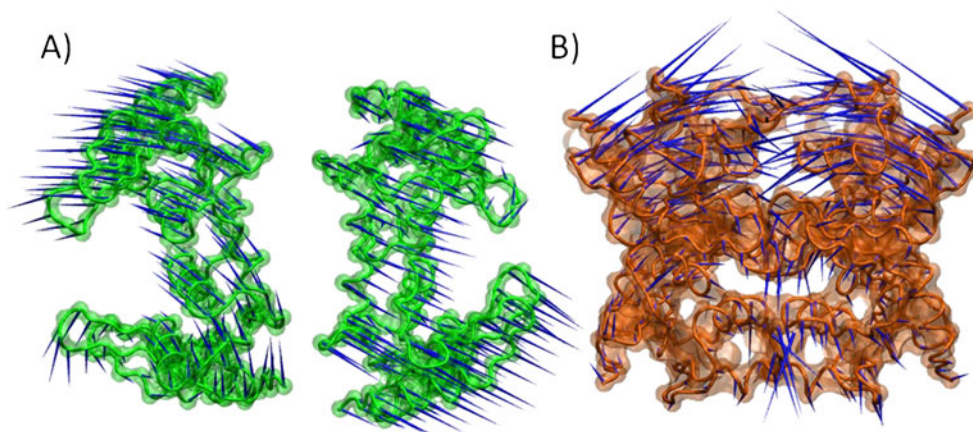## Results and discussion

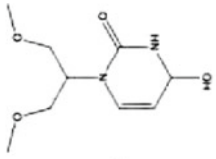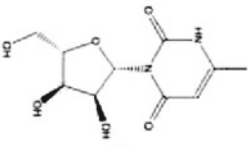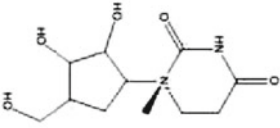Stability of the quaternary structure of the MtCDA

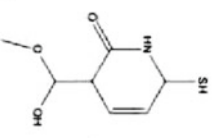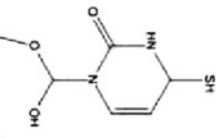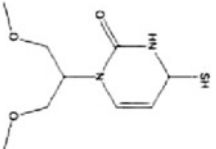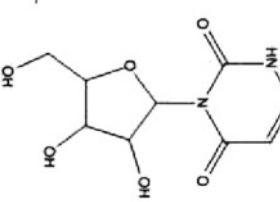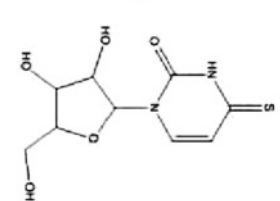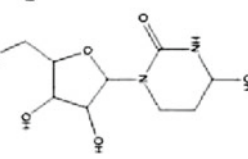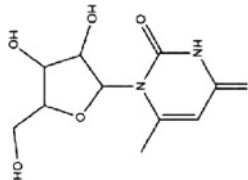In the first work that elucidated the crystal structure of the *MtCDA* [11], it was identified that the quaternary structure was formed by two-fold rotational axes against atomic coordinates of the content of the asymmetric unit. However, alternative quaternary structure is also possible [11]. In order to verify the stability of the structures in solution, we analyzed the root mean square deviation (RMSD) and the radius of gyration (RG) from both structures. Structure A is the quaternary packaging previously identified and structure B is the alternative one. Taking into account the values observed for the RMSD and RG of the structures present in Table 3 and Fig. 2, we can suggest that structure A does not have a reasonable stability in solution when compared to structure B.

Furthermore, analysis of the first eigenvectors represent 69% and 51.9% of the crucial motions during the MD simulations to the structures A and B, respectively. Thereby, the main motions of the structure A are correlated with the dissociation of the quaternary structure, differently from that observed in the structure B (Fig. 3). According to these results, we chose the structure B as the most probable structural packaging for *MtCDA*. A visual inspection of the homotetrameric *CD* from other organisms corroborates with these results, as they present the same quaternary structure found in the present MD simulation. Considering the highest conserved interface regions, which characterize the homotetrameric *CD*, the residues Tyr24, Phe27, Tyr51, Cys59, Arg94, Gln95, Leu119, and Phe123 were analyzed, which correspond to the human *CDA* Phe36, Tyr60, Arg68, Arg103, Gln104, Leu133, and Phe137 residues. These residues represent the most important protein-protein interactions between

Fig. 3 Displacements of PCA model of the structure **a** (previously published) and **b** (proposed structure)

Table 4 The ten best molecules, which were obtained from molecular docking simulations

Id	Structure	Lipinski's rule of five				ΔG*
		HB$_{donor}$	HB$_{acceptor}$	MW	LogP	
Compound 1		1	5	245,32	0,593	-6,2
Compound 2		4	8	260,25	0,394	-9,2
Compound 3		5	8	263,27	-4,98	-8,6
Compound 4		1	4	219,31	0,534	-8,8
Compound 5		1	5	220,29	0,309	-8,8
Compound 6		-	4	261,39	1,546	-6,6
Compound 7		4	8	246,22	-1,16	-8,1
Compound 8		4	7	264,3	-0,19	-8,1
Compound 9		5	8	248,24	-1,14	-7,7
Compound 10		4	5	274,29	-1,3	-8,9

*The values are represented in *kcal*

Fig. 4 The four best molecules encountered by LGA. (**a**) Compound 7, (**b**) Compound 2, (**c**) Compound 5, and (**d**) Compound 3
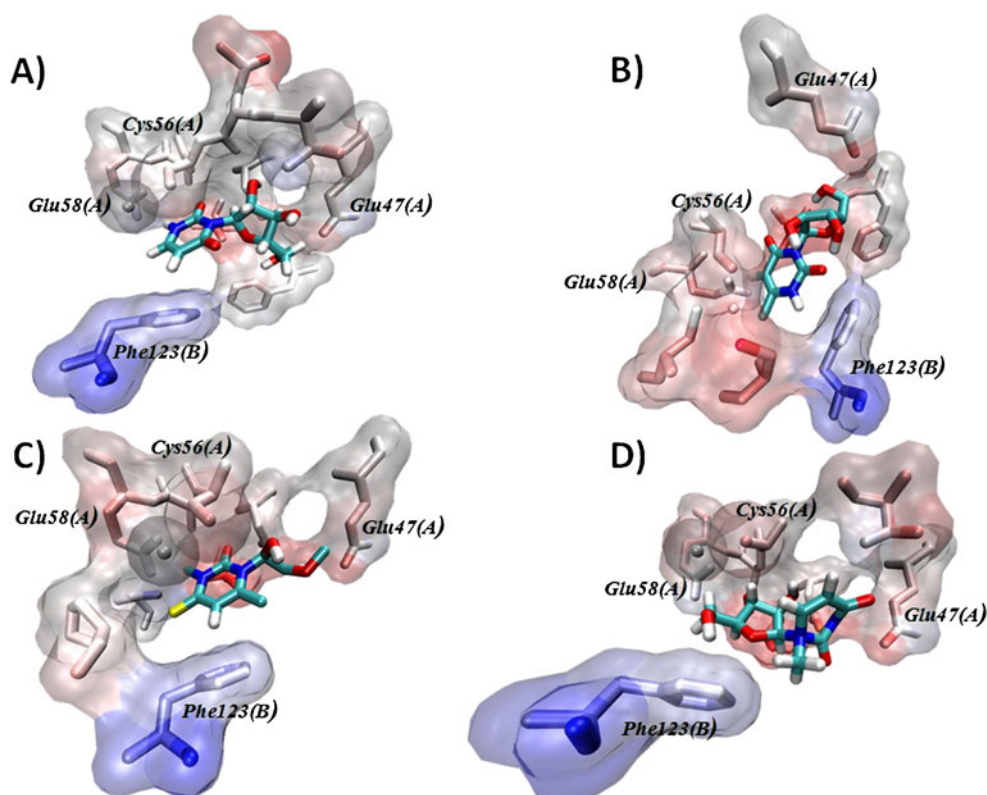
Table 5 The residues making interactions with each of the ten molecules, and the other five crystal structures interacting with their respective ligands

Ligands	HB	vdW	Ligands	HB	vdW	Ligands	HB	vdW
Comp_1		Val29(A)	**Comp_2**		Val29(A)	**Comp_3**		Val29(A)
		Ser50(D)		Phe27(A)	Cys56(A)		Ser50(A)	Asn45(A)
	Asn45(A)	Tyr51(D)		Glu47(A)	Glu58(A)		Thr54(A)	Glu47(A)
	Glu47(A)	Gly52(D)		Ser50(D)	Pro88(A)		Ala57(A)	Tyr51(D)
	Cys56(A)	Ala57(A)		Tyr51(D)	Cys89(A)		Glu58(A)	Cys56(A)
		Glu58(A)		Ala57(A)	Cys92(A)			Phe123(B)
		Cys89(A)			Phe123(B)			
		Phe123(B)						
Comp_4		Val29(A)	**Comp_5**		Val29(A)	**Comp_6**		Val29(A)
		Asn45(A)			Asn45(A)			Glu47(A)
	Cys56(A)	Glu47(A)		Cys56(A)	Glu47(A)		Asn45(A)	Ser50(D)
	Ala57(A)	Ser50(D)		Ala57(A)	Ser50(D)		Gly52(D)	Thr54(A)
		Tyr51(D)			Tyr51(D)		Zn	Leu55(A)
		Glu58(A)			Glu58(A)			Cys56(A)
		Cys78(A)			Cys78(A)			Glu58(A)
		Pro88(A)			Pro88(A)			Phe123(B)
		Phe123(B)			Phe123(B)			
Comp_7	Ser25(A)		**Comp_8**			**Comp_9**		
	Phe27(A)	Val29(A)		Ser25(A)				Ser25(A)
	Glu(47)	Asn45(A)		Phe27(A)	Ser50(D)			Phe27(A)
	Ser50(A)	Cys56(A)		Glu47(A)	Glu58(A)			Val29(A)
	Thr54(A)	Phe123(B)		Val49(A)	Pro88(A)			Asn45(A)

Table 5 (continued)

Ligands	HB	vdW	Ligands	HB	vdW	Ligands	HB	vdW
	Leu55(A)			Cys56(A)	Phe123(B)			Glu47(A)
	Ala57(A)			Cys89(A)				Cys56(A)
	Glu58(A)			Zn				Cys89(A)
								Phe123(B)
1JTK			*1ZAB*			*2FR5*	Asn54(A)	
	Asn42(A)			Asn54(A)	Val38(A)		Tyr60©	Phe36(A)
	Tyr48(A)	Val26(D)		Tyr60(C)	Cys59(.C)		Cys65(A)	Val38(A)
	Ala54(D)	Cys53(D)		Ala66(A)	Glu67(A)		Ala66(A)	Cys59(.C)
	Glu55(D)	Phe125(B)		Cys65(A)	Phe137(B)		Glu67(A)	Phe137(B)
	Cys86(D)						Cys99(A)	
							Cys102(A)	
2FR6	Asn54(A)		*1MQ0*	Asn54(A)				
	Cys65(A)	Val38(A)		Tyr60(B)	Val38(A)			
	Ala66(A)	Phe137(B)		Cys65(A)	Cys59(B)			
	Glu67(A)			Ala66(A)				
	Ser97(A)			Glu67(A)				

Table 6 The main residues which are making interactions with each protein-ligand complexes during the simulation

Inhibitors	Time	Hydrogens bonds	Hydrophobic contacts	
Compound 1	*1ns*		Val22	Glu58
		Asn45	Phe27	Cys89
		Glu47	Val29	Gly124(B)
			Cys56	Leu125(B)
			Ala57	Tyr51(D)
	2ns		Val22	Phe123(B)
			Phe27	Gly124(B)
		Asn45	Val29	Leu125(B)
		Glu47	Cys56	Val49(D)
			Ala57	Ser50(B)
			Glu58	Tyr51(B)
			Cys89	
Compound 2	*1ns*		Phe27	
		Ser25	Val29	
		Asn45	Glu47	
		Thr54	Cys56	
			Tyr51(D)	
	2ns		Phe27	
		Ser25	Val29	
		Asn45	Cys56	
		Thr54	Leu86	
		Tyr51(D)	Asp126(B)	
Compound 3	*1ns*	Glu47	Phe27	Tyr51(D)
		Cys56	Pro88	
		Glu58	Cys92	
		Met87	Phe123(B)	
		Cys89	Ser50(D)	

Table 6 (continued)

Inhibitors	Time	Hydrogens bonds	Hydrophobic contacts			
	2ns		Ser25			
			Phe27			
		Glu58	Val29			
		Leu86	Cys56			
		Arg91	Pro88			
		Cys92	Cys89			
			Phe123(B)			
Compound 4	*1ns*		Phe27	Thr54		
			Val29	Cys56		
			Asn45	Asp124(B)		
			Val46	Leu125(B)		
			Glu47	Tyr51(D)		
	2ns		Ser25	Tyr51		
			Phe27	Thr54		
			Val29	Leu55		
			Val46	Leu125(B)		
			Glu47	Tyr51(D)		
Compound 5	*1ns*		Ser25	Leu125(B)		
		Asp124(B)	Arg26	Leu128(B)		
		Tyr51(D)	Phe27			
			Ala57			
			Phe123(B)			
	2ns		Ser25			
		Asp126(B)	Phe27			
		Tyr51(D)	Ala57			
			Leu125(B)			
			Leu128(B)			
Compound 6	*1ns*		Ser25	Glu58	Leu125(B)	
			Phe27	Leu86	Ser50(D)	
		Ala57	Val29	Met87	Tyr51(D)	
			Glu47		Cys89	
			Cys56		Phe123(B)	
	2ns		Val22	Ser25	Glu58	Leu125(B)
			Phe27	Leu86	Cys92	Ser50(D)
		Ala57	Val29	Met87	Tyr51(D)	Gly127(B)
			Glu47		Cys78	
			Cys56		Phe123(B)	
Compound 7	*1ns*		Phe27	Phe123(B)		
		Arg91	Val29	Gly127(B)		
		Val49(D)	Glu47	Leu128(B)		
			Thr54			
			Cys56			
	2ns	Ser25	Phe27	Gly127(B)		
		Cys56	Val29	Leu128(B)		
		Arg91	Glu47			
		Phe123(B)	Thr54			
		Val49(D)				
Compound 8	*1ns*		Phe27	Leu86	Cys89	Ser50(D)
		Ser25	Val29	Met87	Leu128(B)	

Table 6 (continued)

Inhibitors	Time	Hydrogens bonds	Hydrophobic contacts		
		Asn45	Ala57	Cys78	Pro88
		Glu58	Cys56		Phe123(B)
		Tyr51(D)	Leu125(B)		
	2ns	Ser25	Ala57	Cys78	Pro88
		Glu58	Phe27		Leu86
		Cys87	Val29		Met87
		Tyr51(D)	Cys56		Phe123(B)
			Leu125(B)		
Compound 9	*1ns*		Ser25		Phe27
		Glu58	Val29		Cys56
		Val49(D)	Ala57		Leu86
			Leu128(B)		
			Ser50(D)		
	2ns		Ser25		Phe27
		Ala57	Val29		Cys56
		Cys78	Glu58		
		Ser50(D)	Leu86		
			Tyr51(D)		
Compound 10	*1ns*		Val22		Phe27
		Cys56	Val29		Val79
		Cys78	Asp80		Ser84
		Leu86	Val85		
		Tyr51(D)	Phe123(B)		
	2ns		Val22		Phe27
		Cys56	Val29		Val79
		Cys78	Ala57		Ser84
		Asp80	Leu86		Phe123(B)
		Tyr51(D)	Leu124(B)		

the subunits in order to maintain the quaternary structure. Taking these into account, we observed that Arg68, which is described as an important residue in the catalytic event in the human *CDA* [30], compensating the negative charges of the cysteines, corresponds to a cysteine in the *MtCDA*. Nevertheless, analyzing the *MtCDA* structure, it could be suggested that the residue responsible for counterbalancing the charges of the cysteines is Arg91, which is not conserved in the human *CDA*, where it corresponds to Ala101.

Molecular docking simulation

MtCDA molecular docking simulation was carried out by AutoDock 4.0.2, and all the docking parameters selected are described above. Before the molecular docking process, the partial atomic charges of the ligands were assigned according to the Gasteiger-Hückel [29, 30] method implemented in AutoDock. In order to obtain molecules that have better binding affinities than the substrates and products, we also perform the molecular docking simulation with the substrates and products (cytidine, deoxycytidine, uridine, and deoxyuridine). In this view, only the ligands that presented the estimated free energy of binding $3\,kcal \times mol^{-1}$ lower than substrates and products were considered as potential inhibitors. Besides analysis of the free energy, all 91 compounds were subjected to visual inspection, considering the binding mode and their similarity with the substrates and/or products. Taking into account these filters, ten molecules that may have a potential activity as inhibitors for *MtCDA* were selected. Table 4 presents these molecules, which were ordered by the estimated free energy with AutoDock. When analyzed the Lipinski's rule of five, all ten compounds were in agreement to the main

Table 7 The hit compounds free energy of binding calculated with the LIE equation

Inhibitors	Lennard-Jones^{vdW}	Coulomb^{el}	ΔG_{bind} (Kcal)
Compound 1	-52.71	19.10	-6.27
Compound 2	-52.70	-36.99	-9.67
Compound 3	-40.79	-53.39	-8.99
Compound 4	-55.28	-36.37	-10.00
Compound 5	-43.77	-21.49	-7.48
Compound 6	-59.80	-62.09	-12.20
Compound 7	-38.12	-90.82	-10.88
Compound 8	-48.86	-153.55	-16.20
Compound 9	-45.73	9.27	-5.89
Compound 10	-78.79	-107.11	-17.61

Fig. 5 Inhibitory effect of compound 9

criteria, such as hydrogen bonds, molecular weight, and ClogP. Figure 4 presents the four best molecules encountered by LGA. Interestingly, all ten best molecules, plus the substrates and/or products, presented five main residues making important interactions (hydrophobic or electrostatic): Glu47(A), Cys56(A), Glu58(A), Phe123(B), and Tyr51(D). In addition, the majority of the compounds were coordinated by the zinc ion, similarly to substrates and/or products bound to the binding pocket. By analyzing the distribution of the hydrogen bonds among the hit compounds, we observed that the major contacts are associated in the stability of the ribose moiety. On the other hand, the pyrimidine moiety also presents considerable hydrogen bond interactions, so its importance should not be neglected. Therewithal, the core contribution to the stability of the ligands into the binding cavity is driven by van der Walls interactions, which mostly involve the residues of the adjacent subunits, such as Phe123, Leu125, Asp126, Gly127, and Leu128 of the monomer B, and Val49, Ser50, and Tyr51 of the monomer D. Table 5 shows the main residues which have interactions with each protein-ligand complexes during the simulation.

The presence of the residues from another monomer that can be observed in other tetrameric *CDA*s highlights the importance of the right packaging in order to analyze potential inhibitors. These finding was corroborated when we analyzed other five crystal structures from *Bacillus Subtillis* (1), *Mus Musculus* (3), and human (1) (PDB access codes: 1JTK, 1ZAB, 2FR5, 2FR6, and 1MQ0, respectively) [9, 31], which also presented the residues from other monomers participating in the ligand stability into the active site. Table 6 shows the residues that are making interactions with each of the ten molecules, and the other five crystal structures interacting with their respective ligands.
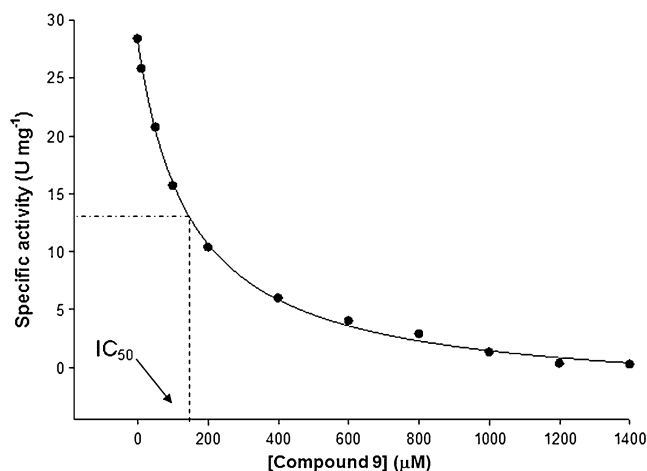
Free energy of binding estimation

In this section, we analyzed the estimated free energy of binding using the LIE method. As described above, 40 simulations were performed (four for each protein-ligand complex) to obtain the Lennard-Jones (a) and Coulomb (b) potentials in order to allow the ΔG of binding calculation. The coefficients were calibrated taking into account five human *CDA* inhibitors with inhibition constants previously described. Prior to calibrating the LIE equation terms, we investigated which kind of interaction was prevalent between the protein-ligand complexes. Finally, the parameters adopted were $\alpha = 0.59$, $\beta = 0.43$, and $\gamma = 0$, as previously mentioned. Table 7 presents the hit compounds free energy of binding. The lowest value was achieved by compound 10 ($-17.61\,kcal \times mol^{-1}$) followed by compounds 8 ($-16.2\,kcal \times mol^{-1}$), 6 ($-12.2\,kcal \times mol^{-1}$), 7 ($-10.88\,kcal \times mol^{-1}$), and 4 ($10\,kcal \times mol^{-1}$). Four of these five compounds (10, 8, 6, and 7) present sulfur group replacing amino or hydroxyl groups, commonly found in the substrate or products, respectively. The zinc ion contribution could be highlighted, mainly by Coulomb interactions, to the inhibitor accommodation into the binding cavity.

$$IC_{50} = 151.74\,\mu M$$
$$\Delta G_{EXP} = -5.2\,kcal \times mol^{-1}$$
$$\Delta G_{LIE} = -5.89\,kcal \times mol^{-1}$$
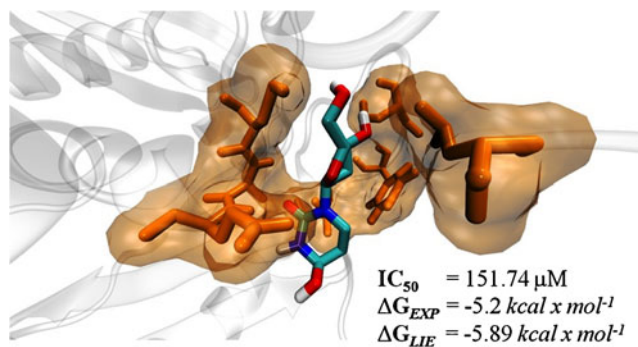
Fig. 6 Compound 9 binding cavity schematic drawing

Compound 4 shows only a hydrogen atom replacing the amino or hydroxyl groups in the pyrimidine moiety, thereby presenting an additional hydroxyl group, which could assist the inhibitor binding mode. On the other hand, the IC_{50} determined for compound 9 ($-5.89\,kcal\times mol^{-1}$), a well-established human CDA inhibitor, was 151.74 μM (~ -5.2 $kcal\times mol^{-1}$) diverging only $0.7\,kcal\times mol^{-1}$ from the value obtained using the LIE equation, demonstrating that the parameters applied into the molecular docking process are capable of finding a potential inhibitor for $MtCDA$.

IC_{50} determination

An IC_{50} value of 151.74 μM was determined for compound 9, a hit compound obtained among several by the molecular docking process (Fig. 5). Although this value might stand at an unacceptable concentration for an ideal inhibitor, it demonstrates a certain level of inhibition, which was not observed for other compounds tested (data not shown) for the same enzyme. Thereby, compound 9 might have the potency improved by chemical derivatization guided by the findings here presented. At any rate, compound 9 represents a hit that, hopefully, could lead to a lead compound with anti-TB activity.

Conclusions

The present study provides structural insight of the $MtCDA$ quaternary structure. Analysis between the region involved in the catalytic event in human CDA ($_{65}$CAERTA$_{70}$) and $MtCDA$ ($_{56}$CAECAV$_{61}$) revealed that the arginine residue found in human CDA, important to compensate the charges in the binding cavity, corresponds to a cysteine in $MtCDA$. Even so, it could be suggested that $MtCDA$ Arg91, corresponding to Ala101 in human CDA, is responsible for counterbalancing the cysteine charges. The molecular docking protocol was capable of identify hit compounds, and suggests that most of them involve interactions among adjacent subunits making clear the importance of van der Walls interactions for ligand binding. Glu47(A), Cys56(A), Glu58(A), Phe123(B), and Tyr51(D) were identified as a core domain to suitably accommodate the ligands into the binding cavity. The IC_{50} value determined for compound 9 (151.74 μM) (Fig. 6) was in agreement with the approach used in the molecular docking process and with the coefficients used in order to estimate the free energy of binding. Furthermore, the coefficients (α, β, and γ) used in view to estimate the free energy of binding through the MD simulations gives promising results, diverging only 0.7 $kcal\times mol^{-1}$ from compound 9 IC_{50} value. Summarizing, our findings open up the possibility to extend this protocol (combining docking and molecular dynamics simulations) to different databases in order to find new potential inhibitors for promising targets based on a rational drug design process.

References

1. Donald PR, van Helden PD (2009) The global burden of tuberculosis–combating drug resistance in difficult times. N Engl J Med 360:2393–2395
2. Frieden TR, Sterling TR, Munsiff SS, Watt CJ, Dye C (2003) Tuberculosis. Lancet 36:887–899
3. Jassal M, Bishai WR (2009) Extensively drug-resistant tuberculosis. Lancet Infect Dis 9:19–30
4. World Health Organization. Communicable Disease Cluster, 1999
5. Ducati RG, Ruffino-Netto A, Basso LA, Santos DS (2006) The resumption of consumption - a review on tuberculosis. Mem Inst Oswaldo Cruz 101:697–714
6. Shambaugh GE (1979) Pyrimidine biosynthesis. Am J Clin Nutr 32:1290–1297
7. Ireton GC, Black ME, Stoddard BL (2003) The 1.14 A crystal structure of yeast cytosine deaminase: evolution of nucleotide salvage enzymes and implications for genetic chemotherapy. Structure 11:961–972
8. Hyde JE (2007) Targeting purine and pyrimidine metabolism in human apicomplexan parasites. Curr Drug Targets 8:31–47
9. Johansson E, Mejlhede N, Neuhard J, Larsen S (2002) Crystal structure of the tetrameric cytidine deaminase from Bacillus subtilis at 2.0 A resolution. Biochemistry 41:2563–2570
10. Wheeler PR (1990) Biosynthesis and scavenging of pyrimidines by pathogenic mycobacteria. J Gen Microbiol 136:189–201
11. Sánchez-Quitian ZA, Schneider CZ, Ducati RG, de Azevedo WF, Jr JCB, Basso LA, Santos DS (2010) Structural and functional analyses of Mycobacterium tuberculosis Rv3315c-encoded metal-dependent homotetrameric cytidine deaminase. J Struct Biol 169:413–423
12. Chung SJ, Fromme JC, Verdine GL (2005) Structure of human cytidine deaminase bound to a potent inhibitor. J Med Chem 48:658–660
13. Vincenzetti S, Mariani PL, Cammertoni N, Polzonetti V, Natalini P, Quadrini B, Volpini R, Vita A (2004) Isoenzymatic forms of human cytidine deaminase. Protein Eng Des Sel 17:871–877
14. Lemaire M, Momparler LF, Raynal NJM, Bernstein ML, Momparler RL (2009) Inhibition of cytidine deaminase by zebularine enhances the antineoplastic action of 5-aza-2'-deoxycytidine. Cancer Chemother Pharmacol 63:411–416
15. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 36:D901–D906
16. Irwin JJ, Shoichet BK (2005) ZINC – a free database of commercially available compounds for virtual screening. J Chem Inf Model 45:177–182

17. 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

18. Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, Reading

19. Solis FJ, Wets RJB (1981) Minimization by Random Search Techniques. Math Oper Res 6:19–30

20. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701–1718

21. 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

22. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular Forces. Reidel, Dordrecht, the Netherlands

23. van Aalten DMF, Bywater B, Findlay JBC, Hendlich M, Hooft RWW, Vriend GJ (1996) PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. J Comput-Aided Mol Des 10:255–262

24. 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 KM, 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, 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, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Head-Gordon M, Replogle ES, Pople JA (2003) Gaussian 03. Gaussian, Wallingford, CT

25. Amadei A, Linssen ABM, Berendsen HJC (1993) Essential dynamics of proteins. Proteins 17:412–425

26. Aqvist J, Medina C, Samuelsson JE (1994) A new method for predicting binding affinity in computer-aided drug design. Protein Eng 7:385–391

27. Vincenzetti S, Quadrini B, Mariani P, De Sanctis G, Cammertoni N, Polzonetti V, Pucciarelli S, Natalini P, Vita A (2008) Modulation of human cytidine deaminase by specific aminoacids involved in the intersubunit interactions. Proteins 70: 144–156

28. Copeland RA (2005) Evaluation of Enzyme Inhibitors in Drug Discovery. Wiley, New York

29. Gasteiger J, Marsili M (1978) A new model for calculating atomic charges in molecules. Tetrahedron Lett 19:3181–3184

30. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital lectronegativity - a rapid access to atomic charges. Tetrahedron 36:3219–3228

31. Teh AH, Kimura M, Yamamoto M, Tanaka N, Yamaguchi I, Kumasaka T (2006) The 1.48 A resolution crystal structure of the homotetrameric cytidine deaminase from mouse. Biochemistry 45:7825–7833

ORIGINAL PAPER

Investigation of the acylation mechanism of class C beta-lactamase: pKa calculation, molecular dynamics simulation and quantum mechanical calculation

Smriti Sharma · Pradipta Bandyopadhyay

Abstract β-Lactamases are bacterial enzymes that act as a bacterial defense system against β-lactam antibiotics. β-Lactamase cleaves the β-lactam ring of the antibiotic by a two step mechanism involving acylation and deacylation steps. Although class C β-lactamases have been investigated extensively, the details of their mechanism of action are not well understood at the molecular level. In this study, we investigated the mechanism of the acylation step of class C β-lactamase using pKa calculations, molecular dynamics (MD) simulations and quantum mechanical (QM) calculations. Serine64 (Ser64) is an active site residue that attacks the β-lactam ring. In this study, we considered three possible scenarios for activation of the nucleophile Ser64, where the activation base is (1) Tyrosine150 (Tyr150), (2) Lysine67 (Lys67), or (3) substrate. From the pKa calculation, we found that Tyr150 and Lys67 are likely to remain in their protonated states in the pre-covalent complex between the enzyme and substrate, although their role as activator would require them to be in the deprotonated state. It was found that the carboxylate group of the substrate remained close to Ser64 for most of the simulation. The energy barrier for hydrogen abstraction from Ser64 by the substrate was calculated quantum mechanically using a large truncated model of the enzyme active site and found to be close to the experimental energy barrier, which suggests that the substrate can initiate the acylation mechanism in class C β-lactamase.

Keywords β-lactamase · pKa calculation · MD simulation · Quantum mechanical calculation · Enzymatic reaction mechanism

Introduction

The production of β-lactamase enzymes is the primary mechanism by which bacteria combat β-lactam antibiotics such as penicillin and cephalosporin, giving rise to antibiotic resistance—a rising healthcare problem. β-Lactam antibiotics act as analogs of the substrates of penicillin-binding-proteins (PBPs). PBPs catalyze the formation of cross-links between the layers of peptidoglycan to form the bacterial cell wall. β-Lactam antibiotics recognize and bind to PBPs covalently. Since PBP cannot hydrolyze the covalently bound β-lactam antibiotics, this leads to bacterial autolysis. β-Lactamase enzymes inactivate β-lactam antibiotics by hydrolyzing the β-lactam ring [1–9]. β-Lactamases have been classified into four classes, A, B, C and D, according to their sequence. Except class B, which is a zinc metallo-enzyme, all other β-lactamases use serine as the nucleophile to cleave the β-lactam ring. In terms of healthcare, the efficacy and prevalence of class C is a major concern, as these are prevalent in Gram-negative hospital-acquired pathogens. They are able to catalyze third generation cephalosporins and are not inhibited by common β-lactamase inhibitors such as clavulanic acid. A major challenge for the scientific community is to design better antibiotics that cannot be cleaved by class C β-lactamase [10, 11]. This requires a very

S. Sharma · P. Bandyopadhyay (✉)
Centre for Computational Biology and Bioinformatics,
School of Computational and Integrative Sciences,
Jawaharlal Nehru University,
New Delhi 110067, India
e-mail: praban07@gmail.com

detailed understanding of the mechanism of action of the enzyme at the molecular level.

There have been many studies of class C β-lactamases using site directed mutagenesis, X-ray crystallography, kinetic experiments and computer simulations [12–18]. The general mechanism of action of class-C β-lactamase is similar to that of serine proteases. In the acylation step, after nucleophilic attack from the enzyme, an acyl-enzyme complex is formed. In the deacylation step, the catalytic water reacts with the acyl-enzyme complex, leading to release of the hydrolyzed product, thereby liberating the free enzyme. The reaction is shown schematically in Fig. 1. Structure A is the pre-covalent complex. Structure B is the acyl-enzyme high-energy intermediate for acylation. Structure C is the acyl-enzyme complex. Structure D is the high-energy intermediate for the deacylation step and, finally, Structure E is the product. In the acylation step, for class C β-lactamases, Serine64 (Ser64) is the nucleophile that attacks the β-lactam ring. However, it is known that the hydroxyl group of serine is not a sufficiently strong nucleophile; therefore, it needs activation from other neighboring amino acid(s), similar to the serine protease mechanism. However, which amino acid(s) actually activates Ser64 is a matter of debate. Based on mutagenesis experiments and X-ray crystal structures, two possibilities must be considered: Tyrosine 150 (Tyr150) and Lysine 67 (Lys67). These residues could act individually or possibly together [19]. There is also a proposal whereby the substrate plays a role in activating Ser64 [20–22]. The important active site residues along with the substrate are shown in Fig. 2 and the possible acylation mechanisms are shown in Fig. 3.

Site-directed mutagenesis experiments involving Tyr150 are so far inconclusive. Dubus et al. [15] reported that Tyr150 shows a substantial loss of activity—as much as 10,000 fold—when studied with various substrates. However, in another study, Dubus et al. [16] showed that the loss of activity is insignificant. One important point is that Tyr150 must be in the deprotonated form to act as an activator, as shown in Fig. 3. This has been investigated both by ^{13}C NMR experiments and pKa calculations [23,

24]. ^{13}C NMR experiments indicated that Tyr150 essentially remains in the protonated form in the free enzyme. This contradicts the Wade groups' pKa calculation, which gave a pKa value of 8.4, suggesting the possibility of a deprotonated form of Tyr150 [23].

In another proposal, Lys67, which is also close to Ser64, has been suggested to act as the general base (Fig. 3: Scheme II). Mutation of Lys67 to glutamine (K67Q) led to a large decrease in activity [14]. A Lys67 to arginine (K67R) mutation also reduced activity, probably by distorting the active site, although the electrostatics may not change significantly. Here again, Lys67 should exist as the neutral form to act as a general base. In an interesting proposal, the Shoichet group [25], based on their crystal structures, suggested that the carboxylate group of β-lactam, which is close to the OH group of Tyr150, can activate Tyr150 by abstracting its proton, and Tyr150 in turn can activate Ser64 (Fig. 3: Scheme IIIA). In an extensive study, Goldberg et al. [13] performed combinatorial scanning mutagenesis to mutate 122 amino acids in class C β-lactamase to ascertain the role of different amino acids as general bases. Their observations indicate that it is difficult to pinpoint a single amino acid as the general base; rather, a collection of amino acids may contribute to make Ser64 a stronger nucleophile.

Crystal structures of wild type and mutant β-lactamase, in both apo and complex forms, have been solved [19, 20, 25, 26]. In a pioneering work, Shoichet et al. [25] determined the structure of the Serine64 to glycine (S64G) mutant along the reaction path. They solved the structure of the pre-covalent complex for the first time (Fig. 1: Structure A). They also solved the structure of the acyl-enzyme complex, and the structure of the product form (Fig. 1: Structures C, E). They noted that the substrate undergoes a large conformational change during the course of the reaction, suggesting the role of the substrate in each step of the reaction. In a recent study, Shoichet et al. [12] investigated the role of Lys67 in acylation and deacylation. Their main conclusion was that a variety of mechanisms may operate in the action of class C β-lactamases depending on the enzyme (wild type or mutant) and the substrate.
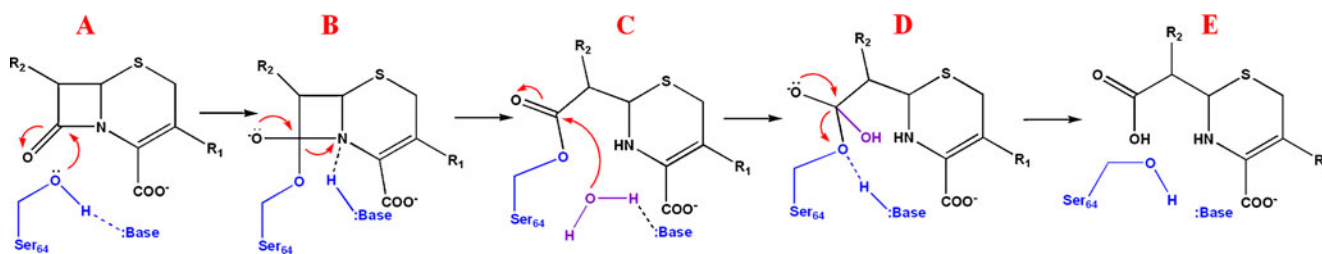

Fig. 1 Schematic reaction mechanism of serine β-lactamases. Reaction proceeds via a pre-covalent complex (Structure *A*), the acyl-enzyme high-energy intermediate (Structure *B*), the acylated enzyme (Structure *C*), the high energy deacylated intermediate (Structure *D*), and the deacylated hydrolyzed product with the free enzyme (Structure *E*)
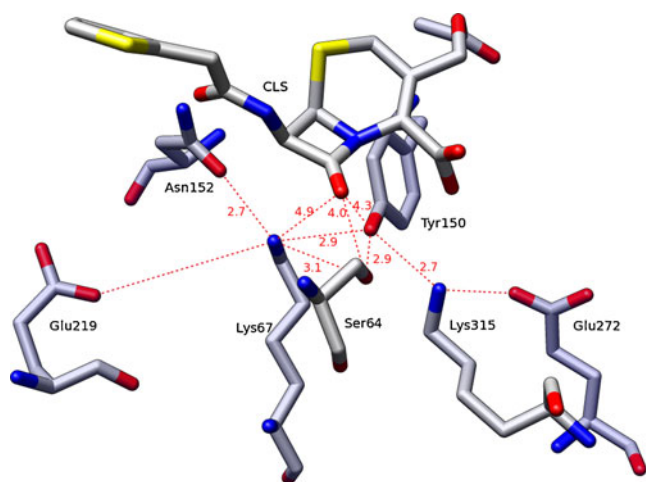
Fig. 2 Hydrogen bonding network in the active site of the enzyme, with the distances (in Å) between key residues as in the pre-covalent complex. Since the pre-covalent complex is a mutant of Serine64 (Ser64), the distance between Tyrosine 150 (Tyr150) and Ser64 is given as its value in the apo enzyme structure 1KE4. Atoms: *light blue* carbon, *red* oxygen, *dark blue* nitrogen, *yellow* sulfur

In a computational investigation of the acylation mechanism of class C β-lactamases, Diaz et al. [27] investigated the stability of different protonation state combinations of Tyr150, Lys67 and Lys315 by performing molecular dynamics (MD) simulations and free energy calculations using a hybrid semi-empirical quantum mechanical/molecular mechanical (QM/MM) energy function. Their observation was that, for the apo enzyme, the free energy difference between the state with neutral Lys67 and protonated Tyr150 and the state with protonated Lys67 and deprotonated Tyr150,

both with the charged state for Lys315, is very small. For the substrate–enzyme complex, the most stable protonation state combination has a neutral Lys67 and a protonated Tyr150. The authors suggested that the role of Lys67 as a general base is probable. In a nice review, Ferrer et al. [22] compared substrate recognition by β-lactamases using different substrates. They concluded that β-lactamases can differentiate between different substrates such as cephalosporin and penicillin, and that the reaction mechanism may differ based on the substrate.

While the number of studies on class-C β-lactamases is quite impressive, conflicting views on the reaction mechanism remain. In the current work, we focused on the acylation step of the reaction mechanism by comprehensive computer simulation, electrostatics calculation and quantum mechanical calculation. We used crystal structures for the apo enzyme [26] and the pre-covalent complex of β-lactamase [25] with cephalothin (CLS) as the substrate. We focused on the role of Lys67, Tyr150 and the substrate, as well as the role of water molecules in the vicinity of the catalytically relevant residues. In the electrostatics calculation, multiple site pKa calculations were performed to include the effect of other titratable residues near the active site. In molecular dynamics (MD) simulations, the geometry of the active site with different combinations of protonation states of Tyr150 and Lys67 was investigated. In pKa calculations, for the apo enzyme, the results show that the population of the protonated form is greater than that of the deprotonated form for Tyr150 and Lys67, although both can exist. And for the pre-covalent complex, it was found that Tyr150 remained essentially

Fig. 3 Schematic diagram for mechanism of acylation. Schemes: *I* Deprotonated Tyr150 acting as a general base to activate Ser64. *II* Neutral Lys67 to act as general base. *III* Substrate activated catalysis: *A* via Tyr150, *B* direct activation
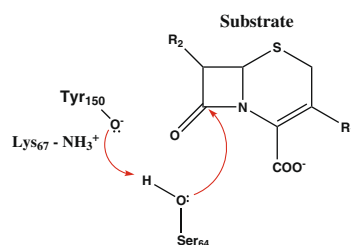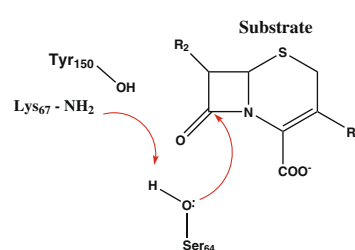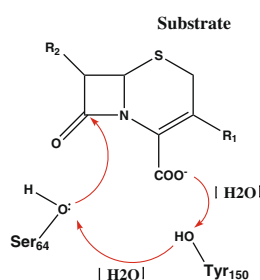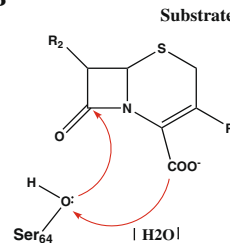
in the protonated form and protonated state population for Lys67 was found to be significantly higher than the deprotonated form. These observations indicate that for CLS as the substrate, Tyr150 and Lys67 alone cannot activate Ser64. For almost 80% of the simulation time, the CO_2^- group of the substrate comes close to the Ser64-OH group. The role of substrate-assisted activation was further explored by quantum mechanical (QM) calculations to estimate the energy barrier for hydrogen abstraction from Ser64 by the substrate using a large truncated model (more than 100 atoms) of the active site. The energy barrier for hydrogen abstraction was found to be comparable to the experimental energy barrier for the acylation reaction of the class C beta-lactamase system. Results from pKa calculations, MD simulations and QM calculations suggest that abstraction of the hydrogen from Ser64 by the substrate could be a viable mechanism to initiate the acylation process.

Materials and methods

pKa calculations

pKa calculations were performed to determine the protonation states of the key titratable residues in the active site of β-lactamase, both in its apo form and in the pre-covalent complex. For the apo enzyme, structures were taken from *Citrobacter freundii* [28] and *Escherichia coli* [26] at 2.0 and 1.7Å resolution (PDB ID 1FR1 and 1KE4, respectively). For the pre-covalent complex, we used a structure obtained from *E. coli* at 1.53Å resolution [25], with substrate cephalothin (CLS) (PDB ID 1KVL). This structure is a mutant with Ser64 mutated to glycine. pKa calculation of the mutant, as well of the wild type, was performed. In the wild type structure, Ser64 was put in place of glycine with the same side chain orientation as in the 1KE4 structure. pKa calculations were also performed on selected snapshots from the MD simulation of the pre-covalent complex (Gly64 was replaced by Ser64 in the simulation). One additional ultrahigh resolution structure, co-crystallized with boronic acid at 1.07Å resolution [19] (PDB ID 2FFY), which has positions of hydrogens, was also studied to check the accuracy of our calculation.

All pKa calculations were carried out with the DELPHI package with the protocol as described by Honig et al. [29]. The pKa value of a residue in a protein was calculated relative to the experimentally determined pKa of a model compound in aqueous solution. The model compound was typically the amino acid with neutral blocking groups, N-methyl (NME) for N-terminal and acetyl group (ACE) for C-terminal. For the single site (residue) pKa calculation, the free energy difference between the protonated and

deprotonated forms for the site in the protein with reference to the same change in the model compound was used to calculate the pKa of the site in the protein. For the multiple site pKa calculation, the charge–charge interactions among the titratable residues contribute to the pKa of a particular site. For N-ionizable residues, the number of states possible is 2^N (since each site can have two states, neutral or charged). The average charge of a site can be determined by statistical mechanical averaging of the charge of that site over all 2^N states. For details, readers are referred to the literature [29–31]. The five conserved titratable residues (Lys67, Tyr150, Glu219, Glu272 and Lys315) located close to Ser64 were considered in the multiple site pKa calculation. To check the effect of other titratable residues, we incorporated two more residues (Tyr112 and His314) in test calculations. However, the results remained almost the same after the addition of these two residues; therefore, they were excluded from the remaining calculations. Hydrogen atoms were added to the crystal structures using the WHATIF software package [32]. The solvent accessible surface (SAS) used to enclose the protein was defined as the volume that could be swept out by a spherical probe of radius 1.4Å. The region outside the SAS was considered as the solvent. Dielectric constants were taken as 80 for the solvent and 4, 10 and 20 for the protein. Amber charges were assigned as the partial atomic charges on the protein and the parse radii for the atomic radii [33, 34]. The deprotonated Tyr and ligand charges were calculated using the standard RESP procedure [35]. To derive the amber charges for the deprotonated Tyr residue, a NME-Tyc-ACE (Tyc referred to the anionic Tyrosine residue) tripeptide was constructed. The RESP module was used to calculate the partial charges derived from the electrostatic potential calculated at HF/6-31 G* level of theory. RESP charges were also derived for boronic acid in 2FFY and for the substrate in 1KVL in the same manner.

Molecular dynamics simulations

Apo enzyme

MD simulations were started using chain A of the 1KE4 structure [26]. The Leap module of the AMBER program package [36] was used to prepare the system for simulation. The *ff99SB* force field [37] with TIP3P [38] water model was used for the simulation. The crystal waters were retained in the simulations. The ionizable residues were set to their normal ionization state at pH 7, except Lys67 and Tyr150, whose protonation states in simulations **I** to **IV** are shown in Table 1. The system was immersed in a water box of size 85×77 x 78 Å³ containing more than 13,000 water molecules. The net positive charge on the system was neutralized through the addition of chloride ions. Bonds

Table 1 Protonation states of Lysine67 (Lys67) and Tyrosine150 (Tyr150) in molecular dynamics (MD) simulations. *CLS* Cephalothin substrate

	Lys67	Tyr150	CLS
Simulation I	Charged	Neutral	NP[a]
Simulation II	Neutral	Neutral	NP
Simulation III	Charged	Charged	NP
Simulation IV	Neutral	Charged	NP
Simulation V	Charged	Neutral	Present
Simulation VI	Neutral	Neutral	Present

[a] Not Present

Fig. 4 Schematic view of the interactions between the active site amino acids in class C β-lactamase and cephalothin (CLS)

involving hydrogen atoms were constrained to their equilibrium position with the SHAKE algorithm.

The system was minimized in two phases. In the first phase, the system was minimized giving restraints (30 kcal mol^{-1} Å^{-2}) to protein and crystallographic waters for 500 steps with subsequent second phase minimization of the whole system. The system was then heated to 300 K over 25 ps with a 1 fs time step in NVT ensemble. The protein atoms were restrained with a force constant of 30 kcal mol^{-1} Å^{-2}. Thereafter, the force constant was reduced by 10 kcal mol^{-1} Å^{-2} in each step to reach the unrestrained structure in three steps of 10 ps each. The system was then switched over to the NPT ensemble and equilibrated without any restraints for 180 ps. The time step for MD simulation for the production run was 2 fs. All four trajectories were run for 10 ns. A non-bonded cutoff of 10Å was used. Constant temperature and pressure conditions in the simulation were achieved by coupling the system to the Berendsen's thermostat and barostat [39]. Periodic boundary conditions were applied to simulate a continuous system. To include the contributions of long-range interactions, the particle-mesh-Ewald (PME) method was used [40]. All MD simulations were carried out using AMBER9 suite of programs. The initial 1 ns of each simulation was discarded from the analysis.

MD simulations of bound form of protein with pre-covalent substrate (cephalothin)

The protonation states of Lys67 and Tyr150 for two pre-covalent simulations (**V** and **VI**) are shown in Table 1. Cephalothin—the bicyclic system of cephalosporins—exists in two conformations, C2-up and C2-down, as this atom of the dihydrothiazine ring can flip between an up- and a down-position with respect to the plane formed by the other atoms of the ring, as shown in Fig. 4. CLS in 1KVL has an extended C2-down conformation. Cephalothin was modeled as a monoanion [41–43] with a charged carboxylic acid attached at C4 of the β-lactam ring. Initial coordinates
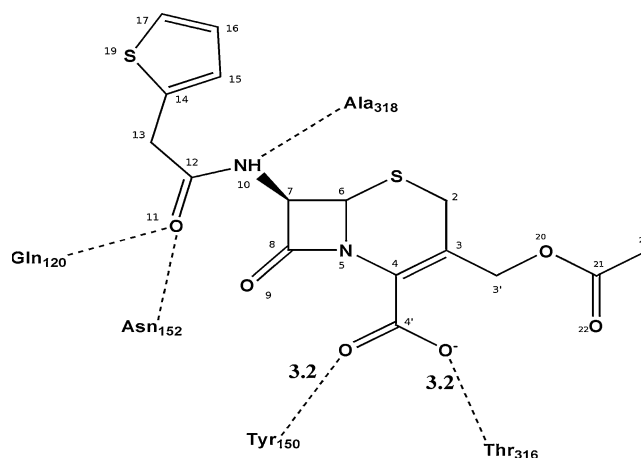
of cephalothin were taken from the crystallographic structure of S64G in complex with cephalothin (PDB ID 1KVL). The atomic charges of CLS were computed using the RESP fitting procedure at the HF/6-31 G* level (charges are given in Supplementary Material). Other force field parameters were taken from AMBER force field using the antechamber program [44]. 1KVL is a S64G mutant structure; therefore, the structure was mutated back to Ser64 using the LEAP module of AMBER. For the complexed enzyme, simulation **V** is the most important, since the pKa calculation indicates that both Lys67 and Tyr150 are essentially protonated (see Results and discussion). Simulation **VI** was performed to check the difference from simulation **V**, as compared to the same change in the apo enzyme. Both simulations **V** and **VI** were run for 10 ns using the same equilibration protocol used for the apo enzyme and the first 1 ns were excluded from the analysis.

Quantum mechanical calculations

Estimation of the reaction barrier of an enzymatic reaction is a major challenge for any computational technique. A QM method must be used to describe the process of bond breaking and forming. However, as QM calculations for the whole enzyme are computationally extremely demanding, this is usually done on a truncated model. The effect of the rest of the system can be accounted for by representing the rest of the protein by a classical potential or a dielectric continuum. For β-lactamase, reaction barrier estimation is quite challenging since the active site contains several charged residues that affect residues outside the active site region. The use of an extensive QM region is required for accurate estimation of the reaction barrier. For the current work, we chose a large truncated model of the enzyme and used full QM calculation. The QM model is composed of all atoms of the three most important residues, Ser64,

Lys67, Tyr150, plus the additional truncated side chain atoms of some of the immediate, hydrogen-bonded residues, i.e., Asn152, Glu272 and Lys315, along with one water molecule. The substrate CLS was taken without its two side chains. The total number of atoms in the calculations was 109. The model is shown in Fig. 5. The truncated regions of the end groups of the critical residues were saturated with linker hydrogen (H). Only the step involving abstraction of the hydrogen from the OH group of Ser64 by the substrate was considered in the current work. Three different representative snapshots were considered in the calculation. The first (Structure I) is close to the crystal structure obtained after minimization of the structure. For the other two snapshots, the distances between the key residues in the active site were determined from the simulation. It was observed that, as the simulation advances, the distances between CLS-CO**O** .... H**O**-Tyr150 and Ser64-**O**H .... H**O**-Tyr150 changed significantly (distances are between the bold atoms). One snapshot (Structure II) was chosen in which these two distances are close to that in the minimized structure, and the other snapshot (Structure III) was chosen in which the distances mentioned are far from the minimized structure distances (distances are given in Supplementary material). Barrier calculation from three snapshots should give a better idea of the closeness of the calculated values to the experimental barrier than calculation from one snapshot only. The reaction coordinate (d1) considered for hydrogen abstraction by the substrate is shown in Fig. 5. The structure was optimized as fixed values of d1 at the RHF/6-31 G* level. During minimization, backbone atoms of the residues were kept fixed. The optimized geometry was then used for single point density functional theory (DFT) calculations using the B3LYP exchange correlation functional with 6-31+G** basis set.
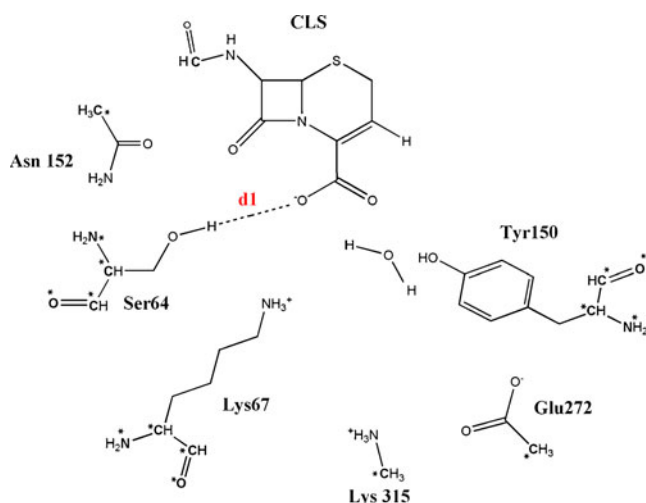
Fig. 5 Model of the enzyme used for quantum mechanical (QM) calculation. Atoms marked with *stars* were frozen during QM optimization

All calculations were done with the GAMESS program package [45]. Two issues regarding our calculation protocol must be mentioned. The first is that whether there could be alternate ways of truncating the protein. In the current protocol, we have made the truncation at the peptide bond. One can also truncate after the peptide bond. To check the differences between these two truncation schemes, reaction barrier calculations for the two truncation models were performed for Structure II. It was found that the energy barriers calculated by these two schemes are within 2 kcal mol^{-1}. This difference is less than the differences in barriers calculated using three different snapshots (see Results and discussion). This justifies the use of our truncation scheme in the calculation. The two reaction profiles are shown in the Supplementary material along with models of the two truncation schemes. The second issue is the use of a continuum model to take care of the rest of the protein. Although this approach has been used by previous workers [46], there are several issues with continuum models especially when one is dealing with systems with several charged sites [47] as is the case here. Moreover, the choice of internal dielectric constant for the active site is very subjective and is used more like a parameter rather than from first principles. Hence, the effect of dielectric continuum was not used in the barrier calculation in the current work, which is based on first principles. However, since the size of our truncated model is relatively large, the effect of the rest of the system on the barrier height is expected to be small.

Results and discussion

pKa calculation

Apo enzyme

Table 2 shows the results of the pKa calculations for the two crystal structures considered for the apo form of the enzyme with a protein internal dielectric constant 20. The pKa values for the 1FR1 structure are found to be slightly higher than that of the 1KE4 structure. pKa values of Lys67 and Tyr150 decrease as the internal dielectric constant decreases. For Lys67, in the 1FR1 structure, the values decrease by only a small margin (9.2 and 9.0 with protein dielectric constants of 10 and 4, respectively). The effect is larger for the 1KE4 structure, where the pKa is 8.7 and 8.3 with internal dielectric constants 10 and 4, respectively. For Tyr150, the values become 9.3 and 9.0 (in 1FR1) and 8.7 and 8.4 (in 1KE4) with internal dielectric constants 10 and 4, respectively. Generally speaking, there is no consensus on the internal dielectric constant to be used in pKa calculations. However, for β-lactamases, previous studies

Table 2 pKa values of three residues in the active site of the apo (PDB ID 1FR1 and 1KE4) and pre-covalent forms (PDB ID 1KVL) of β-lactamase using the continuum electrostatics model with a protein dielectric constant 20. pKa values for the critical residues in the pre-covalent complex (after replacing glycine64 by serine) at different protein dielectric constants are also shown

	Crystal structure		
Fixed protein dialectic constant (=20)	1FR1	1KE4	1KVL
Lys67	9.3	8.9	13.1
Tyr150	9.4	9.0	12.8
Lys315	11.6	10.5	12.3
	Dielectric constant		
Different protein dialectic constants:	4	10	20
Lys67	9.6	9.9	10.2
Tyr150	10.2	10.3	10.3
Lys315	11.1	11.1	11.4

reported values at high internal dielectric constants. The Wade group [23] reported an internal dielectric constant of 78.5 for class C β-lactamases. Oliva et al. [48] reported pKa results for an internal dielectric constant of 20 for TEM-A β-lactamase. The calculated pKa values should be compared with the pKa values for isolated lysine and tyrosine (10.5 and 10.1, respectively) to understand the protonation states of these residues in the protein. It can be concluded that Lys315 should exist in the protonated form because of its high pKa values, which remain essentially the same for different dielectric constants (data not shown). For Lys67 and Tyr150, both protonated and deprotonated states can exist together for the pKa values obtained. However, the population of protonated and deprotonated states depend on the crystal structure taken and internal dielectric constant used.

Pre-covalent complex

The result for the pre-covalent complex for the crystal structure (mutant S64G) is shown in the last column of the upper part of Table 2. The results for the wild type (i.e., after replacing glycine by serine at position 64) are given in the lower part of the Table. It can be seen that the pKa values for Lys67 and Tyr150 (and of Lys315) are higher than 12.0 in the mutant structure. However, it is more important to determine the pKa values of the structure after replacing glycine by serine. Table 2 shows that the pKa values of Lys67 and Tyr150 are about 1 unit higher than in the apo structures. Lys315 again has pKa values greater than 11, suggesting its preference for the protonated form. To understand the pKa values of Lys67 and Tyr150 further, we removed the substrate from the pre-covalent structure of the wild type and then calculated the pKa for this modeled system (data not shown); we found lower pKa values similar to those observed for the apo structures discussed

above. This means that the presence of the substrate causes pKa values to increase. A closer inspection of the structure reveals that the oxygens of the CO_2^- group of the substrate are close to Tyr150 and Lys67. The distance between the oxygen of the CO_2^- group closest to $Tyr150-Oh(Lys67--N_\zeta)$ is 3.2 (5.7) Å. This causes the protonated states of these two residues to be more favorable compared to the apo form. To check if the high pKa results are specific for the pre-covalent crystal structure, and if they are dependent on the orientation of the Ser64 side chain, we investigated further the structures obtained from MD simulation. During MD simulations, the CO_2^- group remained close to Lys67 and Tyr150. pKa calculations of several representative structures obtained from the MD simulation show high pKa values (between 10.3 to 12.1 at dielectric constant 20) for both Lys67 and Tyr150, suggesting a preference for the protonated states of these two residues. Investigation of both MD structures and the crystal structure indicate the favorability of the protonated states of Lys67 and Tyr150 in the pre-covalent complex with CLS. It is known that pKa calculations can be very sensitive to the parameters used and to the structural details. To validate our calculations, one ultra-high resolution crystal structure (PDB ID 2FFY) [19] of class C β-lactamase was taken, where the positions of hydrogens are given. In the 2FFY crystal structure, both Lys67 and Tyr150 are protonated. In our calculation, the pKa values obtained for Lys67 and Tyr150 are ∼9.5 and ∼9.7 for a protein dielectric constant of 20. Here, pKa values did not change much by changing the dielectric constant of the protein. This indicates that Tyr150 is essentially protonated as seen in the crystal structure. For Lys67, the values are slightly lower than one would expect for an unambiguously protonated state, although the protonated population would be significantly greater than the deprotonated population. We can conclude that, although pKa calculations can be sensitive to calculation details, our calculation protocol is giving sensible results for Tyr150 and slightly lower values for Lys67. As the calculated pKa values for Tyr150 in the pre-covalent complex both from MD simulations and crystal structures are above 10.0, it is highly likely that Tyr150 remains essentially in its protonated form in the pre-complexed state. For Lys67, it is also likely that it exists predominantly in the protonated form.

Molecular dynamics simulations

Stability of the trajectories

The stabilities of all six trajectories were monitored by plotting the C_α root-mean-square-deviation (RMSD) values (Fig. 6). It can be seen that all six trajectories have RMSDs
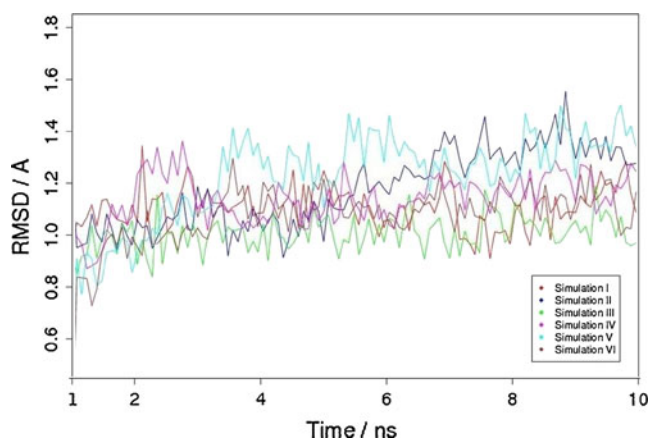
Fig. 6 Time evolution of the root-mean-square-deviation (RMSD) for the six simulations performed

within 1.5 Å during the 10 ns simulation. The RMSDs of the simulations performed with the different protonation states are similar. The RMSD of the pre-complex structure with protonated Lys67 and Tyr150 (Table 1: Simulation **V**) is slightly higher in the initial stage of the simulation. However, in the latter part of the simulation, the RMSD values become similar to that of the apo enzyme.

Structure of the active site

In the X-ray structures, the class C β-lactamase active site is characterized by a dense hydrogen-bonding network that interconnects the catalytically important residues (Fig. 2). In the 1KE4 apo structure, the distance between Oη-Tyr150 and Oγ-Ser64 is 2.9 Å. The N_ζ−Lys67 is 2.8 Å away from Oγ-Ser64. The distance between Tyr150-Oη and Lys67−N_ζ is 3.3 Å (the values are similar in the pre-covalent complex). These three key residues are tightly coupled through hydrogen bonding. The immediate neighbors of these three key residues are Asn152, Lys315 and Thr316. The hydrogen bonds formed by these three residues are also shown in Fig. 2. From the crystal structures, it is clear that both Lys67 and Tyr150 are close

enough to Ser64 to activate it. We next investigated the structural rearrangement of the active site during MD simulations.

Apo enzyme The average distances (and standard deviations) involving the key catalytic residues obtained from the MD simulation are shown in Table 3. The distance between Ser64 and Lys67 (Ser64–CH$_2$Oγ and N_ζ−Lys67) is similar to the crystal structure distance in the first two simulations (~2.9 Å). However, in simulations **III** and **IV** (where Tyr150 is charged), Ser64 and Lys67 move away from each other, increasing the distance to 3.7 and 4.0 Å, respectively, with a corresponding increase in the standard deviations. This indicates that when both Lys67 and Tyr150 are charged or when only Tyr150 is charged, on average Lys67 moves away from Ser64. Thus, geometrically, Lys67 is unlikely to activate Ser64 when Tyr150 is charged. The Ser64–Tyr150 distance increases from its crystal structure distance in all four simulations. In **I** and **II**, where Tyr150 is neutral, this distance is 3.4 and 3.6 Å, respectively. However, the distance increases to 3.9 and 4.2 Å for **III** and **IV**, respectively, when Tyr150 is charged. This probably can be attributed to stronger interactions between the negative charge of Tyr150 and the positive charge of the Lys315, which moves Tyr150 away from Ser64 in simulations **III** and **IV**. The distance between Tyr150-Oη and N_ζ−Lys67 is shortest when both are charged (Simulation **III**), due to strong electrostatic attraction. However, the distance in **I** is shorter than that in **II** and **IV** (where Lys67 is neutral). This is due to the strong attractive force from the positively charged Lys315, which drifts the Tyr150 towards Lys315 in simulations **II** and **IV**, making the Tyr150–Lys67 distance longer. The standard deviation is greater in simulations **II** and **IV,** indicating larger motion of Lys67 and Tyr150 in these cases. The Lys67–Asn152 distance remains almost the same irrespective of the protonation states considered. For the Lys315–Tyr150 distance, the values are about 7 Å and 4.9 Å in simulations **I** and **II**, respectively (where Tyr150 is neutral). For the simulations where Tyr150 is charged (**III** and **IV**), the distances are 2.8 and 3.0 Å (for charged and neutral Lys67), respectively. In short, the distances between

Table 3 Average distances among the key catalytic residues (Ser64-Oγ, Lys67-N_ζ, Tyr150-Oη, Asn152-Oδ1, Lys315-N_ζ) from MD simulations of the apo enzyme and pre-covalent complex. All distances are in Å and standard deviations are given in parenthesis

Simulation No.	I	II	III	IV	V	VI
Ser64_Lys67	2.93 (0.17)	2.92 (0.16)	3.69 (0.78)	3.97 (0.89)	2.87 (0.12)	2.70 (0.09)
Ser64_Tyr150	3.43 (0.33)	3.57 (0.49)	3.86 (0.41)	4.19 (0.41)	3.99 (0.56)	3.09 (0.33)
Lys67_Tyr150	3.25 (0.25)	4.00 (0.78)	2.96 (0.26)	4.67 (0.84)	4.55 (0.50)	4.06 (0.35)
Lys67_Asn152	4.61 (0.31)	4.66 (0.28)	4.91 (0.29)	4.63 (0.60)	4.75 (0.19)	4.77 (0.26)
Tyr150_Lys315	6.98 (0.57)	4.86 (1.92)	2.84 (0.15)	3.04 (0.78)	3.18 (0.55)	3.06 (0.29)

the residues in the active site depend on the electrostatics of the neighboring residues.

Complexed enzyme The average distances between Lys67$-$N$_\zeta$ and O$_\gamma-$Ser64 in simulations **V** and **VI** (2.9 and 2.7Å respectively) remain similar to the apo enzyme simulations **I** and **II** (Table 3). When Lys67 is protonated, the distance between Ser64–CH$_2$Oγ and Tyr150–Oη becomes 4.0Å. The value reduces to 3.1Å when Lys67 is neutral. For apo enzyme, this distance increases for the Lys67 neutral case. This demonstrates the changed electrostatics due to the presence of the substrate. The distance between Tyr150–Oη and N$_\zeta-$Lys67 is more in **V** than in **VI**, although the reverse was observed for same set of simulations for the apo form (**I** and **II**). This is probably due to the loss of interaction between Lys67 and CO$_2^-$ group of CLS in simulation **VI** (where Lys67 is neutral), which moves Lys67 away from both CLS and Tyr150. The Lys315–Tyr150 distance remains almost same during the course of the simulation. The Lys67–Asn152 distance, as in the apo enzyme, does not change significantly for these two simulations, indicating strong coupling between them. Generally speaking, the presence of the CO$_2^-$ group of CLS changes the electrostatics of the active site of β-lactamase compared to its apo form.

Hydrogen bond occupancies

The occupancies of the active site hydrogen bonds are shown in Table 4. The donor–acceptor (D–A) distance threshold was kept at 3.5Å and the threshold for the angle D–H$\cdots$A was kept at 120°. If we compare the results for simulations **I** and **II** with **V** and **VI** (which have the same ionization states of the key residues: **I** and **II** are for apo and **V** and **VI** are for the pre-covalent complex), it can be seen that the presence of the substrate changes the H-bonding pattern for all except for Lys315–Glu272 and Lys67–Tyr150 H-bonds. Of the H-bonds involving the three key residues Tyr150, Lys67 and Ser64, the Ser64–Lys67 H-bond remain intact for simulation **I** and **II** and **V** but not for **VI**. The Lys67–Tyr150 H-bond is not present for any of these four simulations. The Ser64–Tyr150 H-bond is present only in

simulation **VI**. Tyr150–Lys315 H-bonds appear in simulations **V** and **VI**. Based on the above discussion it is clear that the presence of the substrate changes the H-bonding pattern. If we compare simulations **I** and **II** with **III** and **IV**, it can be seen that the major changes occur between Lys315–Tyr150 and Ser64–Lys67. The H-bond occupancy increases for the former while it decreases for the latter. If we sum all the occupancies for each simulation, we can see that occupancy is at least 3.0 and, for the pre-covalent complex, the occupancies are 3.3 and 3.6 for simulations **V** and **VI**, respectively. In summary, there are at least three strong H-bonds keeping the active site in place in all the simulations performed.

Structure of cephalothin

Figure 4 shows the structure of the substrate, CLS. As depicted, there are several residues, such as Gln120, Asn152, Ala318, Tyr150 and Thr316, in the vicinity of CLS that play a key role in holding the substrate firmly inside the active site. During the simulation, side chain movements were observed. The end-to-end distance between S19 and C23 of CLS (shown in the Supplementary material) was monitored and found to be less in **V** than in **VI**. This decreased distance indicates conversion of the extended form of CLS to the compact conformation. The H-bond occupancies between CLS and the residues Asn152 and Ala318 remain at around 60% of the simulation time for simulations **V** and **VI**.

Movement of water and the CLS carboxylate group

From the above discussion on pKa calculation and the geometry of the active site of the enzyme, it can be concluded that, in the presence of the substrate, both the critical residues Tyr150 and Lys67 remain essentially in the protonated form. Lys67 is buried in the protein and no water molecule was found to be close to it in the simulations. Hence, Lys67 cannot be involved in a water-assisted mechanism. One possible mechanism is that Tyr150 can activate Ser64 through a bridging water molecule. Tyr150 may need activation from the CO$_2^-$ group of the substrate. In another possibility, the CO$_2^-$ group of the substrate can activate Ser64 without involving

Table 4 Hydrogen bond (H-bond) occupancies for the key residues (Ser64-Oγ, Lys67-Nζ, Tyr150-Oη, Lys315-Nζ) from MD simulations. A value of 1 implies fully occupied and 0 implies no H-bond formation

Simulation No.	I	II	III	IV	V	VI
Ser64@Oγ … Lys67@Nζ	0.99	0.90	0.14	0.30	0.93	0.00
Ser64@Oγ …Tyr150@Oη	0.00	0.14	0.03	0.02	0.02	0.93
Lys67@Nζ … Tyr150@Oη	0.005	0.007	0.92	0.07	0.04	0.03
Tyr150@Oη … Lys315@Nζ	0.00	0.00	0.97	0.91	0.68	0.64
Lys67@Nζ … Asn152@Oδ1	0.99	0.98	0.85	0.95	0.65	0.99
Lys315@Nζ … Glu272@OE1/2	0.99	0.99	0.48	0.37	0.99	0.99

Tyr150 (Fig. 3: Scheme III). The movement of water molecules in the active site, along with the movement of the CLS carboxylate group, may be critical in understanding the possible mechanisms. Keeping this in mind, the movement of water molecules in the simulation was monitored to check whether there is any water molecule between (1) CLS-CO_2^- and the Tyr150-OH group, or (2) between Tyr150 and Ser64. Also, the movement of CLS CO_2^- was investigated. It was found throughout the trajectory that one water is close to Tyr150 (most of the time the distance is less than 2Å). On the other hand, the distance between the same water and the CLS-CO_2^- group varies between ~2Å to more than 8Å. If a cut-off of 3Å is used, then, for about 40% of the simulation time, the water molecule remains close to both CLS and Tyr150 (see figure in the Supplementary material). The results suggest that activation of Tyr150 by CLS through a water molecule can be a possible route, although it is unlikely to be the major route. For Tyr150-OH and Ser64-OH, the distances between the water molecule closest to both Tyr150 and Ser64 were calculated. It was found that the distance between the hydrogen of the OH group of serine and the oxygen of water varies mostly between 3 and 4Å for most of the simulation. However, the distance between the oxygen of Tyr150-OH and the hydrogen of the water varies widely throughout the simulation, with values ranging from 2 to more than 7Å (fsee figure in the Supplementary material). This indicates that activation by Tyr150 by a bridging water molecule is unlikely. Figure 7 shows the distance between the two oxygens of the CO_2^- group of CLS and hydrogen of Ser64-OH. It can be seen that, except for a small part of the simulation, there is at least one oxygen of CO_2^-, which is close to Ser-OH (less than 2Å). This indicates that CLS-CO_2^- is close enough to activate Ser64 without the help of the intervening water.
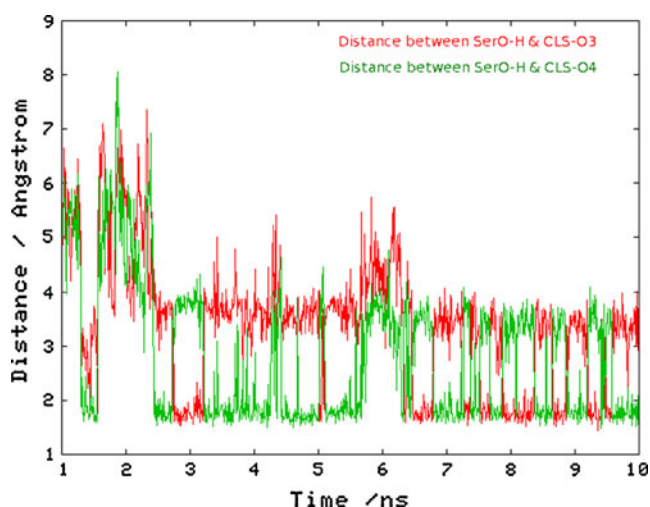
Fig. 7 Distances between the Ser64Oγ-H and oxygens of CO_2^- of CLS are shown for simulation **V**
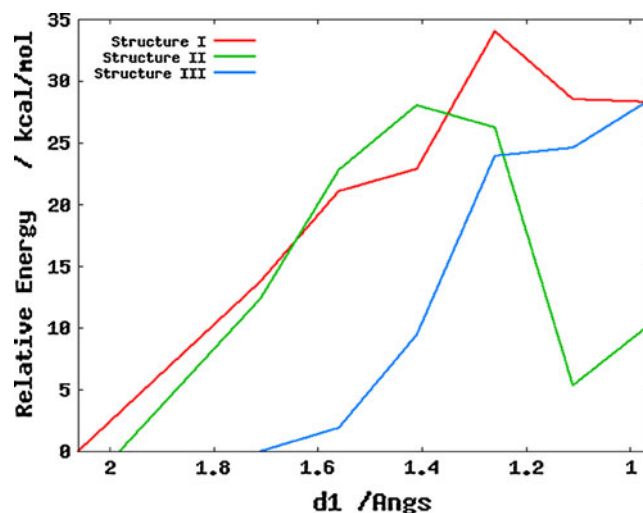
Fig. 8 Quantum mechanical (QM) energy profile for hydrogen abstraction by the substrate for the three snapshots (see text). The energy value at the reaction coordinate value of 1.99 is not shown for structure I (see Table 5)

Quantum mechanical calculations for the abstraction of hydrogen from Ser64 by CLS

Figure 8 shows the energy profile for the one-dimensional reaction coordinate for hydrogen abstraction by the carboxylate group of CLS, and Table 5 depicts the energy values. The three energy profiles represent values obtained from three different starting structures as described in the previous section. No zero point energy (ZPE) correction was added to the energy shown. Also, no entropic contribution was added. As the energy profiles represent only the abstraction of hydrogen from Ser64 by CLS, it cannot be compared directly with the experimental barrier for acylation. However, the values can be taken in order to check if the barrier for the initial step is comparable to the actual barrier. The energy values show that, for structures I and II, the barrier is a few kilocalories per mole higher than

Table 5 Energy barrier of hydrogen abstraction by the substrate in kcal mol^{-1} at the DFT(B3LYP)/6-31+G** level. d1 is the reaction coordinate (in Å). The distances between the hydrogen of Ser-OH and the closest oxygen of carboxylate group of the substrate are 2.06, 1.99 and 1.71 for structures I, II and III, respectively

d1	Structure I	Structure II	Structure III
2.06	0.0		
1.99	−0.3	0.0	
1.71	13.8	12.4	0.0
1.56	21.1	22.8	1.9
1.41	22.9	28.0	9.4
1.26	34.0	26.2	23.9
1.11	28.5	5.3	24.6
0.96	28.3	10.3	28.5

the experimental acylation barrier of 24 kcal mol^{-1} [49]. Since hydrogen abstraction is considered as a one-dimensional reaction coordinate, the estimated barrier is an upper limit to the actual barrier height. Moreover, the backbone of the protein was kept fixed during optimization. It is very likely that without those constraints the barrier will become lower. For structure II, the energy increases at the last point. For the structure III, the energy was found to increase with the reaction coordinate. These results show the difficulties in modeling a complex enzymatic reaction using a single coordinate and with a truncated model. Although the energy profiles calculated are qualitative in nature, it is encouraging that at least two energy profiles have values close to the experimental barrier. The results indicate that the abstraction of hydrogen by CLS is not much higher in energy from the acylation barrier and can be a viable mechanism to initiate the acylation step.

If we combine our results from the pKa calculations, MD simulations and QM calculations to check the mechanisms shown in Fig. 3, schemes I and II are both unlikely, since, in these schemes Tyr150 and Lys67 must be predominantly in deprotonated form, which our electrostatics calculations do not support. This essentially precludes a direct role of Lys67 and Tyr150 in the acylation step, although they may contribute indirectly. The substrate-assisted scheme III shown in Fig. 3 seems most probable since the carboxylate group of the substrate remains close to Ser64 throughout almost the entire simulation, and the energy barrier estimation for the two representative structures gives values comparable to experiments. This indicates that the substrate-assisted mechanism can be a viable tool to initiate acylation. Activation of Tyr150 by CLS through an intervening water molecule can be a competing mechanism, though probably not the main one. The presence of other amino acids is important since they keep the electrostatics and geometry of the active site optimum for activating Ser64. This also supports the view of Goldberg et al. [13], who suggest a role for the H-bonding network in activating Ser64. One issue is whether the substrate-assisted mechanism is universal for all substrates. From our calculations, we can only comment on the mechanism with CLS as the substrate. However, it is likely that the mechanism would depend on the nature of the substrate, since the electrostatics of the active site is very sensitive to details of the substrate structure. One caveat of our MD simulations is that there was no scope for charge transfer during the simulation, which may change the electrostatics dynamically.

Conclusions

The mechanism of action of class C β-lactamases was investigated by performing state-of-the-art electrostatics, MD simulations and QM calculations using several recent crystal structures. We found that, with CLS as the β-lactamase substrate, Tyr150 or Lys67 are unlikely to act alone to activate Ser64 for the substrate, since the probability that they exist in deprotonated states—a necessary condition for activating Ser64—is low. Whether Tyr150 or Lys67 can act in conjunction with a water molecule was also investigated. It was found that Lys67 cannot act in conjunction with a water molecule, as it is almost completely buried in the protein. The water molecule has a low population between Tyr150 and Ser64 during the course of simulation. It seems that the substrate is most likely to activate the nucleophile Ser64, primarily because of the proximity of the CO_2^- group of the substrate to the OH group of Ser64. The QM reaction barrier calculations indicate that, for some representative structures, the barrier is low enough for the substrate-assisted mechanism to be one viable mechanism for acylation.

References

1. Ambler RP (1980) Phil Trans R Soc London B 289:321–331
2. Livermore DM (1995) Clin Microbiol Rev 8:557–584
3. Majiduddin FK, Materon IC, Palzkill TG (2002) Int J Med Microbiol 292:127–137
4. Hall BG, Barlow M (2004) Drug Resist Updat 7:111–123
5. Babic M, Hujer AM, Bonomo RA (2006) Drug Resist Updat 9:142–156
6. Strateva T, Yordanov D (2009) J Med Microbiol 58:1133–1148
7. Bush K, Jacoby GA, Medeiros AA (1995) Antimicrob Agents Chemother 39:1211–1233
8. Pérez-Llarena FJ, Bou G (2009) Curr Med Chem 16:3740–3765
9. Wilke MS, Lovering AL, Strynadka NC (2005) Curr Opin Microbiol 8:525–533
10. Hanson ND (2003) Antimicrob Agents Chemother 52:2–4
11. Philippon A, Arlet G, Jacoby GA (2002) Antimicrob Agents Chemother 46:1–11
12. Chen Y, McReynolds A, Shoichet BK (2009) Protein Sci 18:662–669
13. Goldberg SD, Iannuccilli W, Nguyen T, Ju J, Cornish VW (2003) Protein Sci 12:1633–1645
14. Monnaie D, Dubus A, Frere JM (1994) Biochem J 302:1–4
15. Dubus A, Ledent P, Brasseur JL, Frere JM (1996) Proteins 25:473–485
16. Dubus A, Normark S, Kania M, Page MGP (1994) Biochemistry 33:8577–8586
17. Dubus A, Normark S, Kania M, Page MGP (1996) Biochemistry 34:7757–7764
18. Usher KC, Blaszczak LC, Weston GS, Shoichet BK, Remington SJ (1998) Biochemistry 37:16082–16092
19. Chen Y, Minasov G, Roth TA, Prati F, Shoichet BK (2006) J Am Chem Soc 128:2970–2976

20. Patera A, Blaszczak LC, Shoichet BK (2000) J Am Chem Soc 122:10504–10512
21. Dall'Acqua W, Carter P (2000) Protein Sci 9:1–9
22. Ferrer CF, Frau J, Donoso J, Munoz F (2003) Proteins 51:442–452
23. Lamotte-Brasseur J, Dubus A, Wade RC (2000) Proteins 40:23–28
24. Kato-Toma Y, Iwashita T, Masuda K, Oyama Y, Ishiguro M (2003) Biochem J 371:175–181
25. Beadle BM, Trehan I, Focia PJ, Shoichet BK (2002) Structure 10:413–424
26. Powers RA, Shoichet BK (2005) J Med Chem 45:3222–3234
27. Diaz N, Suarez D, Sordo TL (2006) Biochemistry 45:439–451
28. Oefner C, D'Arcy A, Daly JJ, Gubernator K, Charnas RL, Heinze I, Hubschwerlen C, Winkler FK (1990) Nature 343:284–288
29. Yang AS, Gunner MR, Sampogna R, Sharp K, Honig B (1993) Proteins Struct Funct Genet 15:252–265
30. Nicholls A, Honig B (1991) J Comput Chem 12:435–445
31. Sharp K, Honig B (1990) Annu Rev Biophys Biophys Chem 19:301–332
32. Vriend G (1990) J Mol Graph 8:52–56
33. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179–5197
34. Sitkoff D, Sharp KA, Honig B (1994) J Phys Chem 98:1978–1988
35. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) J Phys Chem 97:10269–10280
36. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2006) AMBER 9. University of California, San Francisco
37. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Proteins 65:712–725
38. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
39. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR (1984) J Chem Phys 81:3684–3690
40. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pederson LG (1995) J Chem Phys 103:8577–8593
41. Diaz N, Suarez D, Merz KM, Sordo TL (2005) J Med Chem 48:780–791
42. Mobashery S, Johnston M (1986) J Biol Chem 261:7879–7887
43. Kuzin AP, Liu H, Kelly JA, Knox JR (1995) Biochem J 34:9532–9540
44. Wang J, Wang W, Kollman PA, Case DA (2006) J Mol Graph Model 25:247–260
45. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su SJ, Windus TL (1993) J Comput Chem 14:1347–1363
46. Hata M, Tanaka Y, Fujii Y, Neya S, Hoshino T (2005) J Phys Chem B 109:16153–16160
47. Yu Z, Jacobson MP, Josovitz J, Rapp CS, Friesner RA (2004) J Phys Chem B 108:6643–6654
48. Oliva M, Dideberg O, Field MJ (2003) Proteins 53:88–100
49. Galleni M, Amicosante G, Frere JM (1988) Biochem J 255:123–129

ORIGINAL PAPER

Study of the interaction between 8-azaguanine and bovine serum albumin using optical spectroscopy and molecular modeling methods

Qiao-Ling Gong · Xin-Gen Hu · Guo-Yong Fang ·
Xin-Hua Li

Abstract The interaction between 8-azaguanine (8-Azan) and bovine serum albumin (BSA) in Tris-HCl buffer solutions at pH 7.4 was investigated by means of fluorescence and ultraviolet-visible (UV-Vis) spectroscopy. At 298 K and 310 K, at a wavelength of excitation (λ_{ex}) of 282 nm, the fluorescence intensity decreased significantly with increasing concentrations of 8-Azan. Fluorescence static quenching was observed for BSA, which was attributed to the formation of a complex between 8-Azan and BSA during the binding reaction. This was illuminated further by the UV-Vis absorption spectra and the decomposition of the fluorescence spectra. The thermodynamic parameters ΔG, ΔH, ΔS were calculated. The results showed that the forces acting between 8-Azan and BSA were typical hydrophobic forces, and that the interaction process was spontaneous. The interaction distance r between 8-Azan and BSA, evaluated according to fluorescence resonance energy transfer theory, suggested that there is a high possibility of energy transfer from BSA to 8-Azan. Theoretical investigations based on homology modeling and molecular docking suggested that binding between 8-Azan and BSA is dominated by hydrophilic forces and hydrogen bonding. The theoretical investigations provided a good structural basis to explain the phenomenon of fluorescence quenching between 8-Azan and BSA.

Keywords 8-Azaguanine · Bovine serum albumin · Fluorescence quenching · UV-Vis spectroscopy · Homology modeling · Molecular docking

Q.-L. Gong · X.-G. Hu (✉) · G.-Y. Fang · X.-H. Li
College of Chemistry and Materials Engineering,
Wenzhou University,
Wenzhou 325035, China
e-mail: hxgwzu@hotmail.com

Introduction

Serum albumin is the most abundant plasma protein in humans and other mammals. It is essential for maintaining the osmotic pressure needed for the proper distribution of body fluids between intravascular compartments and body tissues. It also acts as a plasma carrier by non-specifically binding some hydrophobic steroid hormones and as a transport protein for hemin and fatty acids [1]. Research on the interactions between drug molecules and serum albumin helps to interpret the metabolism and transport process of drugs. The structural and energetic aspects of these processes have been studied in order to provide a rational basis for the fundamental understanding of the interaction and the development of efficient therapeutic agents [2–4].

8-Azaguanine (8-Azan) (Fig. 1) is a purine analog that can inhibit the proliferation of micro-organisms, viruses and tumor cells. It has been studied widely for its biological activity, such as its antineoplastic properties, and has been used in the treatment of acute leukemia. This form of heterocyclic molecule is of particular importance. This is because its structure is composed of a pyrimidinic ring and an azolic moiety, and so many 8-Azan derivatives have been studied in many research fields [5–7]. However, the interaction between 8-Azan and serum albumins has not been reported thus far, so it is of particular interest and importance to explore this interaction from the point of view of pharmacology and pharmacodynamics. In this work, bovine serum albumin (BSA) was studied because of its similarity in structure and its comparability with human serum albumin (HSA) [8–10].

Fluorescence and ultraviolet-visible (UV-Vis) absorption spectroscopy are important methods for determining the interactions between small molecules and bio-
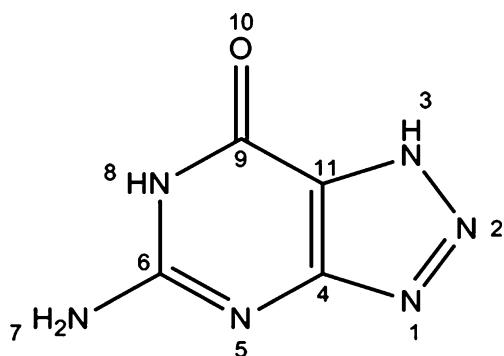
Fig. 1 Molecular structure of 8-azaguanine (8-Azan)

macromolecules. These spectroscopic methods allow non-intrusive measurement of substances at low concentrations under physiological conditions. By analyzing features such as emission peaks and the transfer efficiency of energy, much information can be obtained about the structural changes and the microenvironment of the macromolecule. In this work, fluorescence and UV-Vis absorption spectroscopy were applied to explore the interactions between 8-Azan and BSA. Important information on binding constants, binding sites and thermodynamic parameters was obtained. The effect of energy transfer is also discussed according to the fluorescence resonance energy transfer (FRET) theory. Moreover, theoretical investigations using homology modeling and molecular docking were used to provide a good structural basis for the conclusions drawn.

Materials and methods

BSA and 8-Azan were purchased from Beijing Biodee (Beijing, China) and J & K Scientific (TCI, Tokyo, Japan), respectively. The purity of the BSA is high enough for spectroscopy experiments and prepared solutions, and was thus used with no further purification [11]. The Tris-HCl was purchased from Alfa Aesar (Tianjin, China). All materials, including NaCl and NaOH, were of analytical purity. Tris-HCl buffer solutions ($0.05 \ mol \ L^{-1}$) containing NaCl ($0.05 \ mol \ L^{-1}$) were prepared and the pH was maintained at 7.40. BSA solutions ($5 \times 10^{-6} mol \ L^{-1}$) and 8-Azan solutions ($1 \times 10^{-3} \ mol \ L^{-1}$) were prepared with Tris-HCl buffer solutions (pH 7.40). High-purity water was used in all preparations.

All fluorescence spectra were recorded on an Aminco Bowman Series 2 Luminescence Spectrometer (Thermo Fisher, Waltham,MA), with 1 cm×1 cm×4 cm quartz cuvettes. Fluorescence quenching spectra were performed at 298 K and 310 K in the range of 300–400 nm. The

wavelength of the excitation (λ_{ex}) was 282 nm. The excitation and emission slit widths were set at 8 nm. The UV-Vis spectra were recorded at room temperature on a UV-2501PC (Shimadzu, Japan) spectrophotometer using 1.0 cm quartz cells.

The crystal structure of BSA is not known, and was generated by homology modeling using SWISS-MODEL [12, 13]. The BSA sequence was acquired from the Swiss-Prot protein data bank of the Swiss Institute of Bioinformatics (ID: P02769). Due to the fact that the highest resolution of the crystal structure of HSA is obtained when complexed with heme (PDB: 1N5U, resolution: 0.19 nm), this crystal structure was acquired from the Protein Data Bank and used as the template to make the model of BSA. The WHAT_CHECK program is known to be a good tool with which to assess the value of a homology model [14], but the SWISS-MODEL can analyze the accuracy of a homology model automatically.

Docking between the drug and BSA was performed using Autodock 4.2 [15]. Non-polar hydrogen atoms were added to 8-Azan and BSA and Gasteiger charges were assigned [16]. The $63 \times 63 \times 63$ grid points in the x, y and z dimensions with 0.375 Å spacing were calculated at the most reactive residues. During the docking process, ten conformers were obtained using the Lamarckian genetic algorithm with a small number of evaluations (250,000). Each mode was compared visually using AutoDockTools. The conformer with the lowest binding energy was used for further analysis. PYMOL [17] and Discovery Studio Visualizer 2.5.5 [18] were also used to display the data for the root mean square deviation (RMSD) and to show the crystal structure graphs of homology modeling and molecular docking, respectively.

Results and discussion

Quenching mechanism

Fluorescence spectroscopy is an important tool for studying interactions between molecules because of its high sensitivity and the multiplicity of measurable parameters [19]. Therefore, to investigate the binding of 8-Azan to BSA in detail, fluorescence experiments were carried out. The emission spectra of BSA in the absence and presence of various concentrations of 8-Azan at 298 K and 310 K are shown in Fig. 2. As the data shows, on fixing the excitation wavelength at 282 nm, BSA had a strong fluorescence emission band at 350 nm. The fluorescence intensities of BSA at two different temperatures decreased uniformly, but the emission wavelength of BSA did not change with the increase in 8-Azan
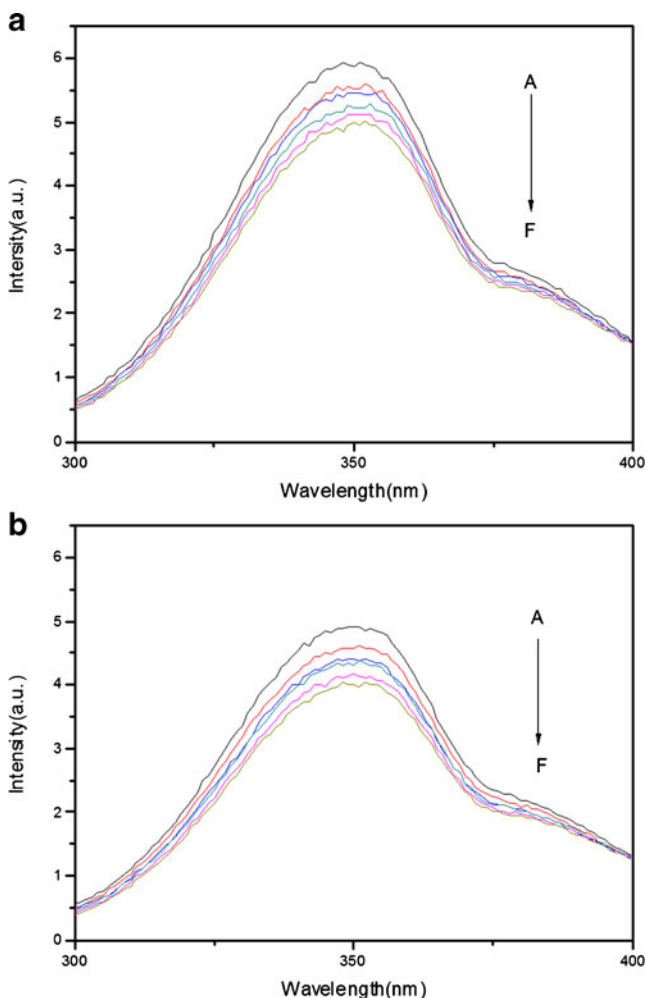
Fig. 2 The fluorescence quenching spectra of bovine serum albumin (BSA) in the presence of 8-azaguanine (8-Azan) at λ_{ex}=282 nm and pH=7.40 (T_a=298 K; T_b=310 K); [BSA] = 5×10^{-6} mol L^{-1}; [8-Azan] (×10^{-6} mol L^{-1}) = 0, 2, 4, 6, 8, 10, respectively, from A to F

concentration, which indicated that an 8-Azan-BSA complex had maybe formed that led to quenching of the fluorescence of BSA.

Quenching occurs by different mechanisms that are usually classified as dynamic quenching or static quenching. Dynamic and static quenching can be distinguished by their differing dependence on temperature and viscosity. Dynamic quenching is collisional quenching, which is enhanced by increasing temperature, but diminished by increasing viscosity. A higher temperature results in faster diffusion and therefore a larger amount of collisional quenching. Static quenching refers to the interaction that the fluorophore makes with the quencher to form a stable non-fluorescent complex. A higher temperature typically results in the dissociation of the weakly bound complexes, and hence a smaller amount of static quenching.

For dynamic quenching, the mechanism can be described by the Stern-Volmer equation [20, 21]:

$$\frac{F_0}{F} = 1 + K_q\tau_0Q = 1 + K_{sv}Q \qquad (1)$$

where F and F_0 represent the fluorescence intensities with and without quencher, respectively; K_q is the bimolecular quenching rate constant; K_{sv} is the dynamic quenching constant; τ_0 is the molecular average lifetime without the quencher (the fluorescence lifetime of the biopolymer is 10^{-8} s) [22, 23] and [Q] is the concentration of the quencher.

With regard to the fluorescence quenching mechanism of BSA by 8-Azan, it is hypothesized that the interaction proceeds via a dynamic route. The corresponding dynamic quenching constants for the interaction between 8-Azan and BSA can be calculated and are given in Table 1. The Stern-Volmer plot of fluorescence quenching of BSA by 8-Azan at 298 K and 310 K can be seen in Fig. 3. For dynamic quenching, the maximum scatter collision quenching constants (K_q) of various quenchers with the biopolymer are 2×10^{10}L mol^{-1} s^{-1} [24, 25]. But, in the present study, the quenching rate constants of BSA initiated by 8-Azan are much greater than this value at both temperatures, which indicates that the probable fluorescence quenching mechanism of BSA by 8-Azan is static quenching, and that an 8-Azan–BSA complex has been formed.

To further investigate the phenomenon, the UV-Vis absorbance spectra and the decomposition of fluorescence spectra were analyzed (Figs. 4, 5). As shown in Fig. 4, the absorption peaks of BSA and 8-Azan were 277.5 nm and 278 nm, respectively, while with the gradual addition of 8-Azan, the maximum absorption spectra showed no significant change. This indicates that static quenching is involved mainly because dynamic quenching affects only the excited state of the quenching molecule, while it has no effect on the absorption spectrum of quenching substances. In Fig. 5, the emission peak is separated into two peaks using a Gaussian function. It can be seen from Fig. 5 that the maximum wavelengths of the two peaks were observed at 345.5 nm and 363 nm, respectively, both of which were very close to the two peaks of BSA and 8-Azan (350 nm and 358 nm, respectively). It is reasonable to consider that they originate from BSA and 8-Azan, respectively, and that BSA also has a small blue shift from 350 nm to 345.5 nm. These results also indicate that there is a static quenching interaction between 8-Azan and BSA. To sum up, it was concluded that the mechanism involved was static quenching.

Table 1 Stern-Volmer quenching constants and the binding parameters for the interaction between 8-azaguanine (8-Azan) and bovine serum albumin (BSA) at T=298 K and 310 K

T(K)	Quenching constants			Binding parameters		
	$10^4 K_{sv}$(L mol^{-1})	$10^{12} K_q$(L mol^{-1} s^{-1})	R	$10^2 K_a$(L mol^{-1})	n	R
298	1.59	1.59	0.9983	6.03	0.71	0.9920
310	1.90	1.90	0.9896	9.21	0.73	0.9883

R Correlation coefficient

Binding constants and sites

For static quenching, when small molecules bind independently to a set of equivalent sites on a macromolecule, the apparent binding constant K_a and binding sites n can be described by the following equation [26]:

$$\lg \frac{F_0 - F}{F} = \lg K_a + n \lg Q \tag{2}$$

where F_0 and F are the fluorescence intensities before and after the addition of the quencher, and [Q] is the total quencher concentration. Figure 6 is the plot of log $(F_0-F)/F$ versus log [Q] for the 8-Azan–BSA system.

By fitting Eq. 2, the apparent binding constant (K_a) and the number of binding sites (n) can be calculated (Table 1). From the binding parameters, it can be seen that the values for the binding site at 298 K and 310 K were nearly equal, and the binding constants K_a were higher at 310 K than that at 298 K. These results indicate that there is a higher binding affinity and a much more stable compound between 8-Azan and BSA at increased temperature.

Thermodynamic parameters

The molecular forces between protein interactions and a drug may include van der Waals interactions, hydrogen bonds, electrostatic interactions and hydrophobic forces. Thermodynamic parameters such as the enthalpy change (ΔH) and entropy change (ΔS) are important for confirming the binding mode, which can be calculated with Eqs. 3 and 4 [27]:

$$\ln \frac{(K_a)_2}{(K_a)_1} = \frac{\Delta H}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \tag{3}$$

$$\Delta G = -RT \ln K_a = \Delta H - T\Delta S \tag{4}$$

where K_a are the binding constants at the experimental temperatures T_1 and T_2, respectively, and R is the gas constant. The free energy change (ΔG) can be calculated with the binding constants from Eq. 4, and, if the temperature does not change a lot, the enthalpy change (ΔH) can be regarded as a constant. In the interaction of 8-Azan with BSA, binding studies were carried out at 298 K

Fig. 3 Stern-Volmer plots of fluorescence quenching of BSA in the presence of 8-Azan at T=298 K and 310 K
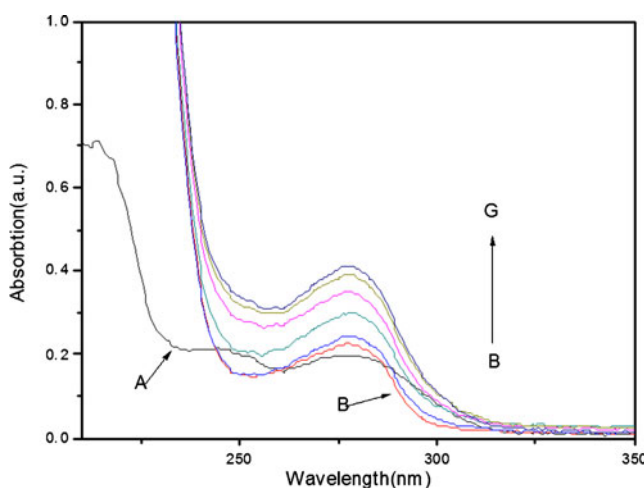
Fig. 4 Ultraviolet-visible (UV-Vis) absorption spectra of the 8-Azan–BSA system. *A* Absorption spectrum of 8-Azan only, [8-Azan] =5.0× 10^{-6} mol L^{-1}; *B–G* [BSA]=5.0×10^{-6} mol L^{-1}: 0, 2, 4, 6, 8, 10 (×10^{-6} mol L^{-1}), respectively, of [8-Azan]
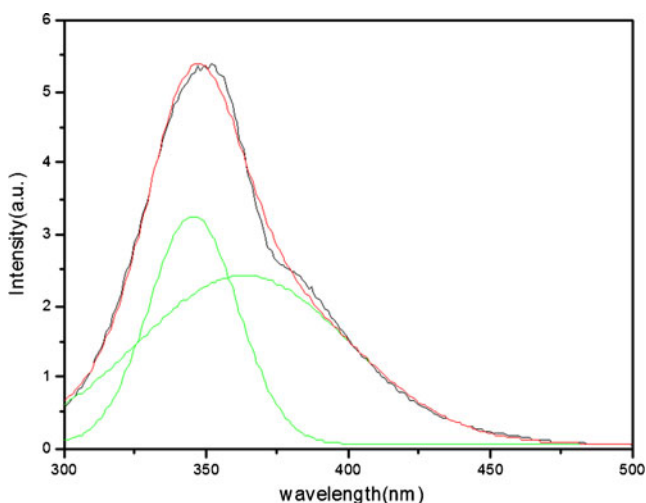
Fig. 5 Decomposition of fluorescence spectra of the 8-Azan–BSA complex excited at 282 nm and T=298 K; [BSA]=5.0×10^{-6} mol L^{-1}; [8-Azan]=5.0×10^{-6} mol L^{-1}

and 310 K, and BSA did not exhibit any structural degradation. The thermodynamic parameters calculated are shown in Table 2.

The negative values of ΔG and its more negative change with the rise in temperature imply that the interaction process is spontaneous and that a high temperature favors a spontaneous interaction. The positive values of ΔH and ΔS indicate that the typical hydrophobic forces play a major role in the binding process.

Energy transfer

FRET is a distance-dependent interaction between different electronic excited states of molecules in which excitation energy transfers from one molecule (donor) to another molecule (acceptor) without the emission of a photon from the former molecular system [28]. According to Förster's
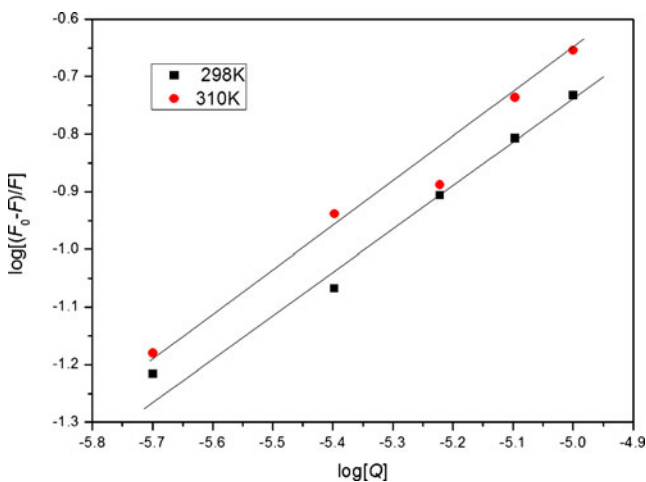
Fig. 6 Plot of log $(F_0-F)/F$ vs log $[Q]$

non-radioactive energy transfer theory [29], the efficiency of FRET depends mainly on the following factors: (1) the donor emission and the acceptor absorption overlap more; (2) the fluorescence produced by the donor has a sufficiently long lifetime; and (3) the distance between the donor and the acceptor. The efficiency of energy transfer between the donor and the acceptor (E) is given by the following equation [9]:

$$E = 1 - \frac{F}{F_0} = \frac{R_0^6}{R_0^6 + r^6} \tag{5}$$

where r is the distance between the donor and the acceptor and R_0 is the distance at 50% transfer efficiency.

$$R_0^6 = 8.79 \times 10^{-25} K^2 n^{-4} \phi J \tag{6}$$

where K is the orientation factor related to the geometry of the donor and acceptor of dipoles, n is the refractive index of the medium, ϕ is the fluorescence quantum yield of the donor, and J expresses the degree of spectral overlap between the donor emission and the acceptor absorption (Fig. 7c), which is given by

$$J = \frac{\int_0^\infty F(\lambda)\varepsilon(\lambda)\lambda^4 d\lambda}{\int_0^\infty F(\lambda)d\lambda} \tag{7}$$

where $F(\lambda)$ is the fluorescence intensity of the fluorescent reagent when the wavelength is λ and $\varepsilon(\lambda)$ is the molar absorbance coefficient at wavelength λ.

In the present case, K^2=2/3, n=1.36 and ϕ=0.15 [30]. From Eqs. 5–7, J, E, and R_0 can be calculated, and thus r can also be calculated, and these values are summarized in Table 2. Because $r<8$ nm, and at the same time, $0.5R_0<r<1.5R_0$, it was concluded that there is a high possibility that energy transfer from BSA to 8-Azan occurs and that the quenching mechanism is of the static type.

Homology modeling

The crystal structure of BSA was built using SWISS-MODEL. The result shows that BSA shares strong sequence similarity with HSA, and their sequence identity is 75.77%. The structural alignment result between the modeler structure of BSA and the crystal structure of HSA is shown in Fig. 8. The RMSD over all C_α-atoms is 0.55 Å, which illustrates that the model of BSA is reasonable. A Ramachandran plot was used to assess the quality of the model structure [31, 32].

Figure 9 shows that approximately 92.9% of the residues were located in the most favored regions, 6.2% in additionally allowed regions and only 0.9% of residues in generously allowed regions. Because a good quality model was expected to have over 90% in the most favored

Table 2 Thermodynamic parameters and energy transfer parameters of the 8-Azan–BSA system

T(K)	Thermodynamic parameter			Energy transfer parameter			
	ΔH(kJ mol^{-1})	ΔG(kJ mol^{-1})	ΔS(J mol^{-1} K^{-1})	J (cm^3 L mol^{-1})	E	R_0 (nm)	r (nm)
298	27.6	−15.86	144.0	3.08×10^{-15}	0.09	1.94	2.84
310		−17.59		3.10×10^{-15}	0.11	1.94	2.75

regions [32], this phenomenon also indicates that the model structure is reasonable and suitable for further studies.

BSA is composed of 607 amino acids and its three-dimensional structure is similar to a heart shape. It contains three structurally similar α-helices domains, I (His27-Gln219), II (Arg220-Ile411) and III (Lys412-Ala607), and each domain contains two sub-domains, respectively. Domains I and II or domains II and III are linked by the stretched helixes that form the two longest helical chains. The BSA molecule has two tryptophan residues: Trp-237, which is in sub-domain IIA, located within a hydrophobic binding pocket of the protein, and Trp-158, which is in the first sub-domain IB of the albumin molecule, located on the surface of the albumin molecule [33].

Molecular docking

The most probable binding position of 8-Azan to BSA using molecular docking is shown in Fig. 10. The figure reveals that the docking region of 8-Azan binding to BSA is located in the hydrophilic cavities in the lowest part of sub-domain IB. The best docked conformation of 8-Azan and BSA, in which various residues are selected within 4Å of the ligand, is shown in Fig. 11. It can be seen that there are hydrogen interactions between N5 of 8-Azan and the residue Asp132, NH$_2$7 and Asp132, NH8 and Ser 216, NH$_3$ and Gln219 in the 8-Azan-BSA system, and the lengths of the hydrogen bonds are 2.16Å, 1.86Å, 2.34Å and 2.46Å, respectively. Moreover, Asp132, Ser 216 and Gln219 are hydrophilic residues. These results suggested that binding between 8-Azan and BSA is dominated by hydrophilic force and hydrogen bonding [34]. However, the results of docking are not very close to those of thermodynamic studies. A reasonable explanation may be that the crystal structure of BSA differs from that of the solution system used in this
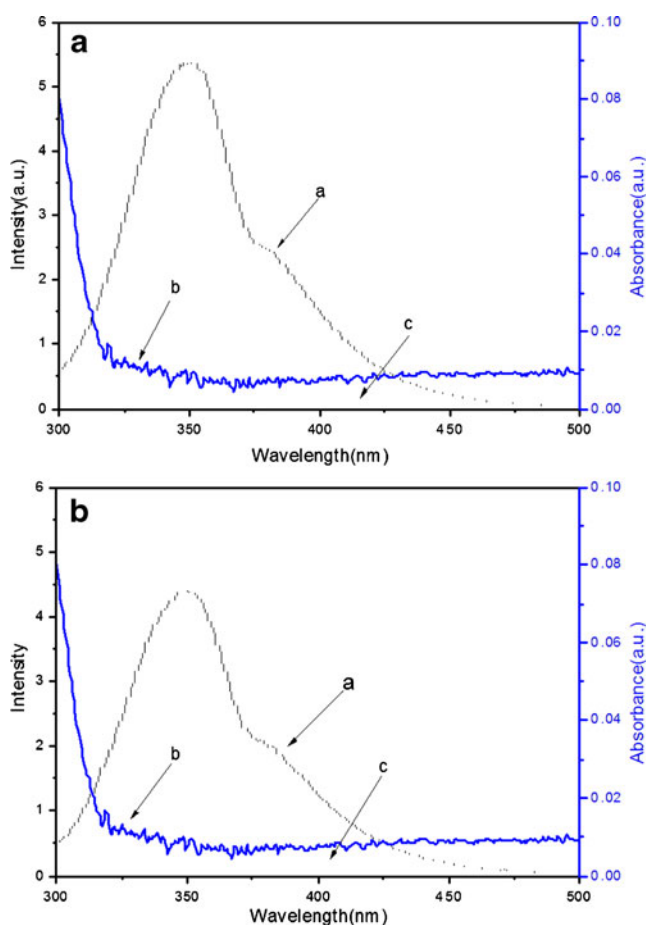
Fig. 7 Fluorescence spectra of BSA (**a**) and the absorption spectra of 8-Azan (**b**); [BSA] =5×10^{-6} mol L^{-1}; [8-Azan] =5.0×10^{-6} mol L^{-1}; T_a=298 K; T_b=310 K

Fig. 8 Superposition models of BSA (*green*) and human serum albumin (HSA; *red*)
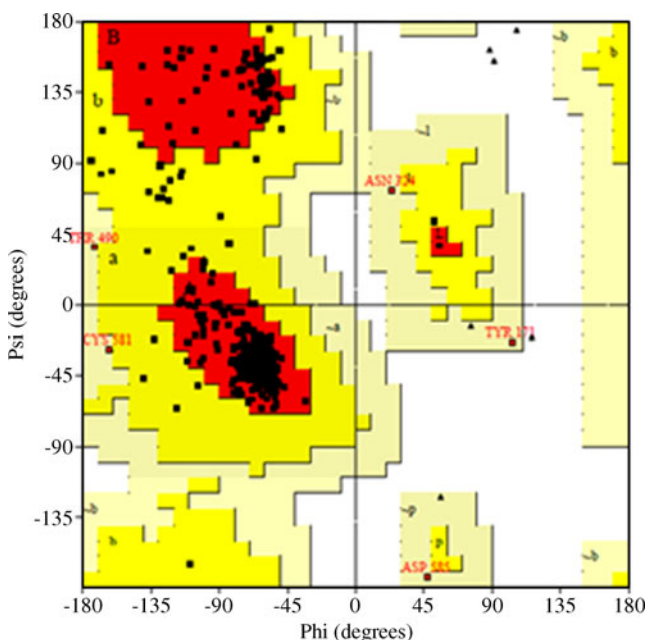
Fig. 9 Ramachandran plot. *Red* Most favored regions [*A, B, L*]; *yellow* additional allowed regions [*a, b, l, p*]; *light yellow* generously allowed regions [*~a, ~b, ~l, ~p*]; *other areas* disallowed regions

Fig. 11 Predicted binding mode between 8-Azan and BSA. The residues of the albumin are represented using *lines* and the 8-Azan structure is represented using a *ball and stick* model. The hydrogen bonds between 8-Azan and the albumin are represented using *green dashed lines*

study. In conclusion, molecular docking provides a good structural basis to explain the phenomenon of fluorescence quenching between 8-Azan and BSA.

Conclusions

The interaction between 8-Azan and BSA in Tris-HCl buffer solutions was investigated using fluorescence and

Fig. 10 Structure of 8-Azan–BSA complex

UV-Vis spectroscopy at 298 K and 310 K, at a wavelength of excitation λ_{ex}=282 nm. From the fluorescence intensity changes and the Stern-Volmer equation, it was found that the main quenching mechanism between 8-Azan and BSA was static quenching. This conclusion was supported by the analysis of UV-Vis absorbance spectra and the decomposition of the fluorescence spectra. In addition, the binding constants (K_a) and the number of binding sites (n) were calculated from the static quenching equation. These results indicate that there is a higher binding affinity and a much more stable compound between 8-Azan and BSA at increased temperature. Furthermore, the negative sign of Gibbs free energy shows that the interaction process is spontaneous. The positive entropy and enthalpy change indicates that the forces acting between 8-Azan and BSA are typically hydrophobic. Meanwhile, according to FRET theory, the distances (r) between the donor (8-Azan) and the acceptor (BSA) are 2.84 nm and 2.75 nm at 298 K and 310 K, respectively, which suggests that there is a high possibility of energy transfer from BSA to 8-Azan occurring by a static quenching mechanism.

Furthermore, the 3D structure of BSA was generated successfully by homology modeling, and was found to be reasonable and suitable for further study. Molecular docking indicated that the interaction between 8-Azan and BSA is dominated by hydrophilic forces and hydrogen bonding, which does not reflect the findings of thermodynamic studies. The reason may be that the crystal structure of BSA is different in different environments.

References

1. Zunszain PA, Ghuman J, Komatsu T, Tsuchida E, Curry S (2003) BMC Struct Biol 3:6
2. Tang D, Li HJ, Li P, Wen XD, Qian ZM (2008) Chem Pharm Bull 56:360–365
3. Kandagal PB, Seetharamappa J, Ashoka S, Shaikh SMT, Manjunatha DH (2006) Int J Biol Macromol 39:234–239
4. Il'ichev YV, Perry JL, Simon JD (2002) J Phys Chem B 106:452–459
5. Zeng XH, Deng SH, Wang HM, Zheng AH, Chen P (2010) Acta Cryst E66:o3140–o3141
6. Lu SF (2004) Microchem J 77:37–42
7. Da Costa CP, Fedor MJ, Scott LG (2007) J Am Chem Soc 129:3426–3432
8. Mahammed A, Gray HB, Weaver JJ, Sorasaenee K, Gross Z (2004) Bioconj Chem 15:738–746
9. Cao SH, Wang DD, Tan XY, Chen JW (2009) J Solution Chem 38:1193–1202
10. Qi ZD, Zhou B, Xiao Q, Shi C, Liu Y, Dai J (2008) J Photochem Photobiol A 193:81–88
11. Heng Chiat Tai (2004) X-ray crystallographic studies of bovine serum albumin and *Helicobacter pylori* thioredoxin-2. PhD thesis, University of Saskatchewan
12. Arnold K, Bordoli L, Kopp J, Schwede T (2006) Bioinformatics 22:195–201
13. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009) Nucleic Acids Res 37:387–392
14. Hooft RWW, Vriend G, Sander C, Abola EE (1996) Nature 381:272–272
15. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) J Comput Chem 30:2785–2791
16. Gasteiger J, Marsili M (1980) Tetrahedron 36:3219–3228
17. DeLano WL (2002) The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA
18. Discovery Studio Visualizer 2.5.5 (2010) Accelrys Inc, San Diego
19. Wang YP, Wei YL, Dong C (2006) J Photochem Photobiol A 177:6–11
20. Hu YJ, Liu Y, Zhang LX, Zhao RM, Qu SS (2005) J Mol Struct 750:174–178
21. Lehrer SS (1971) Biochemistry 10:3254–3263
22. Wang T, Xiang BR, Wang Y, Chen CY, Dong Y, Fang HS, Wang M (2008) Colloids Surf B 65:113–119
23. Lakowica JR, Weber G (1973) Biochemistry 12:4154–4170
24. Wu YH (2007) J Pharm Biomed Anal 44:796–801
25. Hao SZ, Liu SD, Wang XH, Cui XJ, Guo LP (2009) J Lumin 129:1320–1325
26. Alain M, Michel B, Michel D (1986) J Chem Educ 63:365–366
27. Bi SY, Song DQ, Tian Y, Zhou X, Liu ZY, Zhang HQ (2005) Spectrochim Acta A 61:629–636
28. Hu YJ, Liu Y, Jiang W, Zhao RM, Qu SS (2005) J Photochem Photobiol B 80:235–242
29. Kandagal PB, Kalanur SS, Manjunatha DH, Seetharamappa J (2008) J Pharm Biomed Anal 47:260–267
30. Zhang YZ, Zhou B, Liu YX, Zhou CX, Ding XL, Liu Y (2008) J Fluoresc 18:109–118
31. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Cryst 26:283–291
32. Patel A, Dewangan R, Khatri S, Choubey J, Gupta SK, Verma MK (2009) J Eng Technol Res 1:039–045
33. Wang YQ, Zhang HM, Zhang GC, Tao WH, Fei ZH, Liu ZT (2007) J Pharm Biomed Anal 43:1869–1875
34. Cattoni DI, Kaufman SB, González Flecha FL (2009) Biochim Biophys Acta 1794:1700–1708

In silico investigations of possible routes of assembly of ORF 3a from SARS-CoV

Hao-Jen Hsu · Wolfgang B. Fischer

Abstract ORF 3a of human severe acute respiratory syndrome corona virus (SARS-CoV) has been identified as a 274 amino acid membrane protein. When expressed in Xenopus oocytes the protein forms channels. Based on bioinformatics approaches the topology has been identified to include three transmembrane domains (TMDs). Since structural models from experiments are still lacking, computational methods can be challenged to generate such models. In this study, a 'sequential approach' for the assembly is proposed in which the individual TMDs are assembled one by one. This protocol is compared with a concerted protocol in which all TMDs are assembled simultaneously. The role of the loops between the TMDs during assembly of the monomers into a bundle is investigated. Molecular dynamics simulations for 20 ns are performed as a short equilibration to assess the bundle stability in a lipid environment. The results suggest that bundles are likely with the second TMD facing the putative pore. All the putative bundles show water molecules trapped within the lumen of the pore with only occasional events of complete crossing.

Keywords Docking · Ion channel · Membrane protein · Molecular dynamics · Protein assembly · 3a of SARS-CoV

H.-J. Hsu · W. B. Fischer (✉)
Institute of Biophotonics, School of Biomedical
Science and Engineering, National Yang-Ming University,
Taipei, Taiwan
e-mail: wfischer@ym.edu.tw

Introduction

SARS-CoV has a single positive strand RNA genome carrying 14 open reading frames (ORFs), encoding viral structural proteins (such as spike, envelope, membrane, and nucleocapsid proteins), replicases, and accessory proteins [1]. ORF 3a of SARS-CoV is identified as a 274 amino acid (a.a.) structural protein, which is located between S and E proteins [2]. ORF 3a protein harbors three transmembrane domains (TMDs) at its N-terminal side and a longer intracellular C-terminal region of about 148 amino acids. The central region of 3a protein consists of cysteine-rich domain (a.a. 127–133), Yxxφ domain (a.a. 160–163) and diacidic domain (a.a. 171–173) [1–3]. 3a protein is suggested to form a homotetramer via monomer disulfide bridges (Cys-133 [4]) forming a dimer and the noncovalent assembly of two of the dimers forming the functional tetramer [5]. Structural information about the protein or its biological role in the cellular life cycle of the virus is still in the dark.

Viral channel forming proteins, have also been found for other viruses [6–8], such as M2 from influenza A [9–12], Vpu from HIV-1 [13, 14], 8a from SARS-CoV [5, 15], protein p7 from HCV [16, 17], 2B from Polio virus [18, 19] 3a and E proteins from SARS-CoV [4, 20], just to mention some of them, are also known to homo-oligomerize. The number of TMDs increases going from M2, Vpu and 8a, with a single TMD per monomer, to two TMDs for p7 and 2B and finally to three TMDs for 3a. Albeit the emergence of more and more structural information derived from experiments (for a review see [7, 21]) modeling the assembly of the proteins is still a challenge.

In general, a two-staged mechanism for helix-bundle membrane protein folding is suggested [22]: (i) the fold of the membrane domain into its secondary structure, a helix,

(ii) the assembly of the inserted helices in the membrane. This model is expanded to include a third stage in which co-factor insertion, folding of the extramembrane parts and quaternary assembly is included [23]. This model description also holds on the energy landscape for membrane protein folding [24]. Along the line of this mechanism some computational assembly protocols are designed by using rigid body movements to explore energy landscapes [25–27]. These methods allow improving the quality of sampling conformational space via the step width of structure placements. Another approach reported in the literature includes extended replica exchange molecular dynamics simulations to assembly homooligomers [28]. Another almost unbiased approach is achieved if helices can freely diffuse within the lipid bilayer. This approach has been demonstrated for the assembly of TMDs into dimers using coarse-grained MD simulations techniques [29]. Still the full story of membrane protein folding remains to be elucidated [30].

Previous work proposed an assembly methodology to search the conformational space of all possible assemblies for the preferable structure, taking symmetry considerations into account [31]. The methodology has been tested on M2 from influenza A showing agreement with experimentally derived structure. The structure of protein 3a from SARS-CoV was first time predicted based on this methodology. Although it's good agreement with the experiments, the mechanism of simultaneous assembly is still hard to imagine. One idea proposed herein is that there are two assembly methods, concerted and sequential [32], for comparison to search the most preferable bundle models of 3a from SARS-CoV. Loops between the transmembrane domains (TMDs) are also predicted for comparison. MD simulations of possible bundle models are performed for confirmation.

Computational methods

Ideal helices, of the TMDs of 3a [31], $TMD1_{39-59}$ (AS^{40} LPFGWLVIGV^{50} AFLAVFQSA), $TMD2_{79-99}$ (FI^{80} CNLLLLFVTI^{90} YSHLLLVAA), and $TMD3_{105-125}$ (FLY LYA^{110} LIYFLQCINA^{120} CRIIM), were generated with backbone dihedrals of $\varphi = -65°$ and $\phi = -39°$ using the program MOE (Molecular Operation Environment, www. chemcomp.com) and its integrated protein builder.

Equilibration of the TMDs

Each of the ideal helix was embedded into a fully hydrated POPC lipid bilayer (16:0−18:1 diester PC, 1-palmitory-2-oleoyl-*sn*-glycero-3-phospho-chloine) for 10 ns MD simulation to derive a relaxation of the conformation. POPC topology parameters were taken from [33]. And the bilayer

has undergone a 70 ns MD simulation to be equilibrates as much as possible [34]. After insertion a stepwise energy minimization and equilibration protocol was adapted [34]. The system was heated gradually to 310 K in 500 ps, and then five stages of equilibration were performed where all the heavy atoms of the bundle were restrained in their initial positions by applying a harmonic force in x, y and z directions (1000, 500, 250, 100, and 10 kJ/mol/nm). These runs were to adjust the lipid to the inserted bundle.

Prior to assembly

A principal component analysis (PCA) over the backbone atoms of all frames of the last 3 ns of each of the TMDs has been done. A structure was calculated averaging over the first few eigenvectors. PCA was accomplished using the program g_covar from the GROMACS-4.0.5 package. Rotational and translational motions were removed by fitting the peptide structure of each time frame to the starting structure.

Assembly

The equilibrated TMDs derived from MD simulations are used to generate tetrameric assemblies via various routes (Fig. 1).

Monomer assembly

(1) Sequential assembly

The helical backbone structure from PCA analysis is aligned along the z-axis. Two methods were used to assemble the monomer (Fig. 1): sequential 1 (Seq1, assembly from C-terminus to N-terminus) and sequential 2 (Seq2, assembly from N-terminus to C-terminus). For Seq1, TMD3 and TMD2 were assembled first becoming a new TMD unit (TMD3 + TMD2). This unit is consequently assembled with TMD1 to form a monomeric subunit ((TMD3 + TMD2) + TMD1). Herein, one of the TMDs was fixed and the second TMD was rotated around the other TMD (rotational angle 2) and around its own helical axis (rotational angle 1), then tilted and translated to the other TMD. The same rotation protocol was adopted for the generated unit of two TMDs being kept fixed whilst the third TMD was rotated around the unit. Similarly, for Seq2, TM1 is first assembled with TM2 to form a new TMD unit followed by the assembly with TMD3 to form a monomer ((TMD1 + TMD2) + TMD3).

(2) Simultaneous assembly

In the simultaneous assembly three TMDs are assembled in a concerted fashion to form a monomeric subunit [31]. This assembly method is hither forth

503
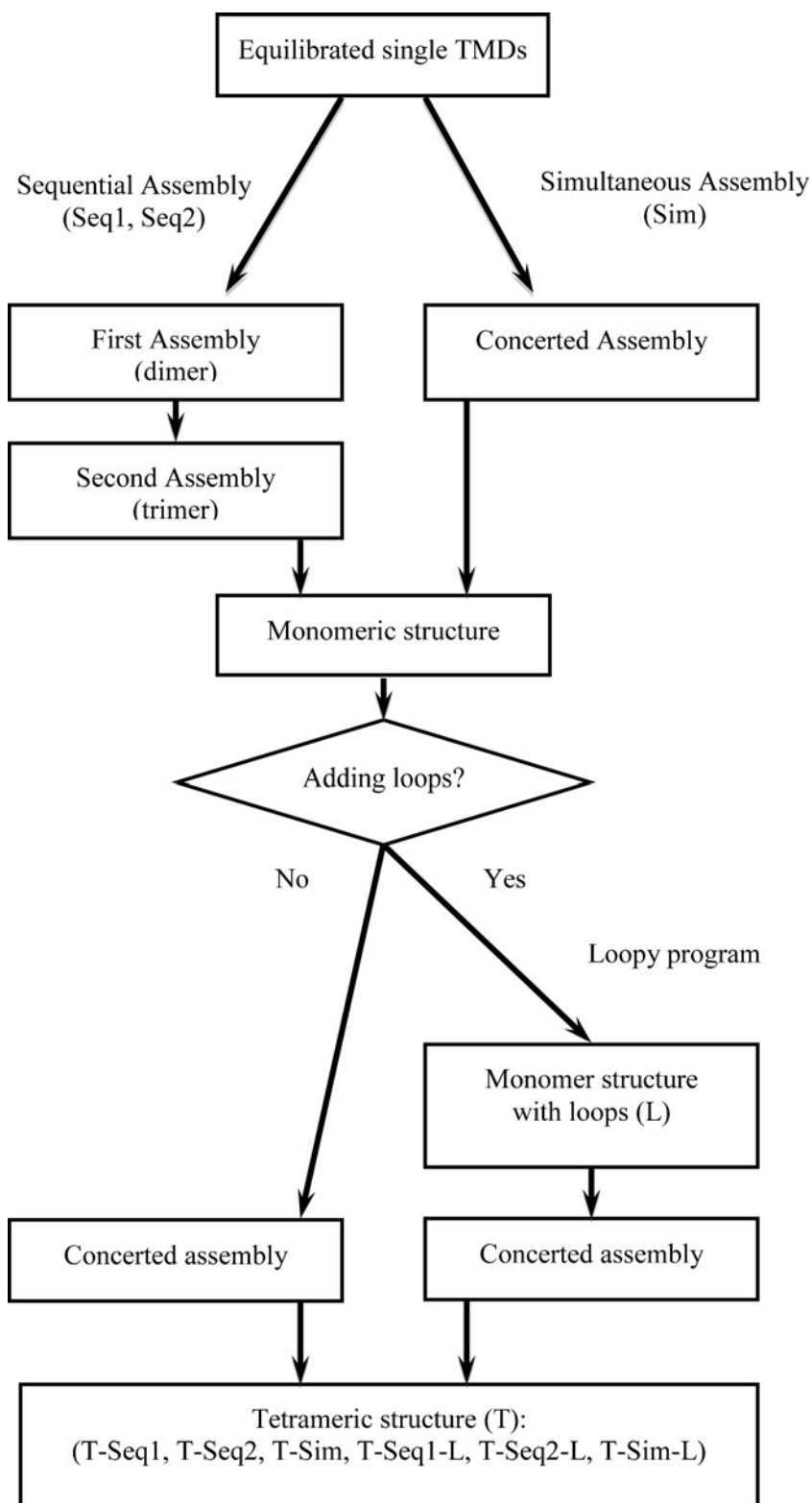
referred to as Sim. According to the symmetry all single TMD backbones were rotated around their own helical axis in the same sense with respect to the central pore axis, and were also tilted simultaneously. The construction of a trimer followed basic geometry with inter-helical separation angles of 120°. Besides, to cover all weak and tight packing inter-helical distances in the range from 8.5 to 12.0Å were sampled for each monomer assembly method. The distance data are referred to as the distance between the center of mass of each TMD.

(3) Adding loops

Monomeric models 'with loops' obtained their 'loops' using the program Loopy [35, 36]. Two loops (loop1: residues 60–78; loop2: residues 100–104) were added on the monomeric subunit accordingly. The lowest energy structures are named Seq1-L, Seq2-L, Sim-L.

Tetramer assembly

The monomeric subunits were assembled into a tetrameric bundle using the Sim protocol. According to the protocol used for the monomeric subunit the tetrameric bundles are referred to as T-Seq1, T-Seq2 and T-Sim (with added loops T-Seq1-L, T-Seq2-L and T-Sim-L). The interhelical separation angle was set to 90° and the interhelical distances sampled in the range of 18 to 24 Å to cover all possible packing modes.

To further sample conformational space, several degrees of freedom were varied systematically, such as interhelical distance by 0.25 Å, rotational angle by 5°, and tilt angle (hither forth called tilt) by 2°. After each positioning, side chain atoms were reconstructed, followed by an energy minimization of 5 steps of steepest descend and 10 steps of conjugated gradient. Potential energy of each conformation/position was evaluated based on the Amber 94 force field in an implicit lipid environment characterized by a dielectric constant of $\varepsilon = 2$. With this protocol hundreds of thousands of different conformations were generated.

MD simulations

The selected tetrameric bundle was then embedded into a POPC lipid bilayer system by removing overlapping lipids and waters molecules. After energy minimization, 4 or 16 Cl⁻ ions were added to compensate for the positive net charge of each monomer. Finally, the whole system without adding loops to the bundle consisted of the bundle (2624 atoms), 462 POPC- and 14616 SPC-water molecules including 4 Cl⁻ (70500 atoms in total). The system with added loops consisted of the 3640 bundle atoms, 462 POPC-, and 14604 SPC-water molecules, including 16 Cl⁻ (71492 atoms in total). The MD simulation protocol was as followed, after energy minimization (see above), 20 ns production runs were carried out without any constraint on the bundle.

GROMACS-4.0.5 with the Gromos96 (ffG45a3) force field was used for the simulations. The simulations were conducted in the *NPT* ensemble employing the velocity-rescaling thermostat at constant temperature 310 K, and 1 bar. The temperature of the protein, lipid and the solvent were separately coupled with a coupling time of 0.1 ps.

Semi-isotropic pressure coupling was applied with a coupling time of 0.1 ps and a compressibility of 4.5 x 10^{-5} bar⁻¹ for the xy-plane as well as for the z-direction. Long range electrostatics calculated using the particle-mesh Ewald (PME) summation algorithm with grid dimensions of 0.12 nm. Lennard-Jones and short-range Coulomb interactions were cut off at 1.4 and 0.8 nm, respectively.

The simulations were run on a DELL Precision T5400 workstation, and a cluster consisting of 32 cores (Xeon 2.26 GHz). Plots and pictures were generated using xmgrace, VMD and MOE.

Results

Equilibration

All three TMDs show stable root mean square deviation (RMSD) values over the entire duration of the simulation. Within the last 4 ns of the 10 ns simulation values between 0.1 and 0.2 nm are calculated identifying that the short run deliver reasonably equilibrated structures (Fig. 2a). The root mean square fluctuation (RMSF) of the individual residues of the TMDs shown in Fig. 2b is indicative for low dynamic of the amino acids with higher fluctuation at either end of the TMDs. The residues of the core region of the TMD, albeit at very low level, exhibit slightly higher dynamics than the residues in the head group region (appr. residues 10 – 25 and 55 – 70) giving the graph a w-like shape. A sequence of residues from Asn-82 to Leu-85 and around Leu-94 to Leu-96 of TMD2 shows a localized area of larger RMSF values.

Assembly

Generation of the monomer

Analysis of the energetic of the monomer assemblies (Suppl. Fig. 1) reveals mostly some close clustering of lowest values independent of the sequence used. Using the first part of Seq1, assembling TMD2 and TMD3, reveals a dimer with lowest energy of −503.6 kcal mol⁻¹ (Table 1 and Suppl. Fig. 1a). Assembling the third TMD, TMD3, results in two low energy structures calculated with values of −878.1 kcal mol⁻¹ and −865.12 kcal mol⁻¹ and interhelical distances of 1.2 and 1.15 nm, respectively (Suppl. Fig. 1b). Both structures are separated by their individual rotational angles but adopt the same tilt direction of −2° and −10°, respectively.

Seq2 reveals a dimer of TMD1 and TMD2 of −349.6 kcal mol⁻¹. Adding TMD3 results in a monomer of −863.5 kcal mol⁻¹, with an interhelical distance of 1.175 nm.
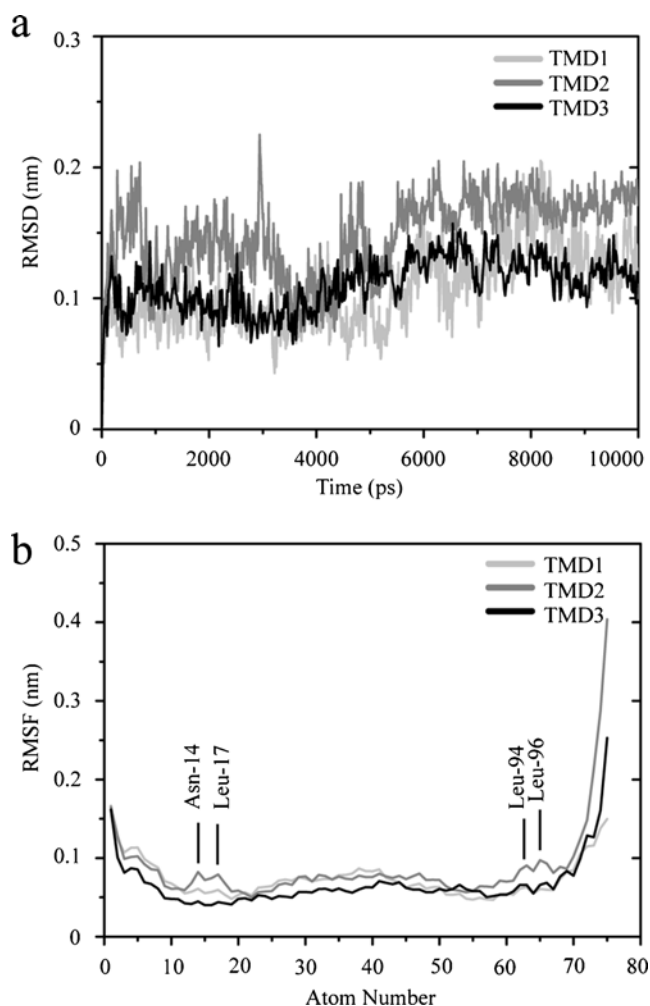
Fig. 2 Root mean square deviation (RMSD) of the Cα backbones of the single TMDs, TMD1, TMD2 and TMD3, referring to the respective starting structure (**a**). Root mean square fluctuation (RMSF) of the atoms of the amino acids (**b**). The TMDs are overlaid so that the atom numbers match for each TMD. Values for TMD1 are shown in light gray, those for TMD2 in gray and the values for TMD3 in black

Finally for Sim the lowest energy structure (−730.5 kcal mol^{-1}) is clearly distinct from the second lowest (−654.8 kcal mol^{-1}) by 75.7 kcal mol^{-1} (data not shown). For these two structures the rotation difference of TMD2 is of up to 40° and with an opposite tilt direction (−4°) than the second lowest (+4°). Similar to the monomer Seq1 the lowest energy structure has a hydrophobic pore and therefore the second lowest monomer is considered further. The interhelical distance of the second lowest structure is 1.075 nm.

The monomeric structure from Seq1 (Fig. 3a) exhibits a hydrophilic stripe, which is due to residues of TM3 (Tyr-109, Tyr-113, Gln-116, and Asn-119) spanning the entire TM stretch. Residues like His-93, Tyr-89 and Asn-82 of TMD2 join toward the same direction. The monomeric structure assembled from Seq2 indicates two hydrophilic stripes, one from Ser-92, His-93, Thr-89, and Asn-82 (all TMD2), and the other from Tyr-109, Tyr-113, Gln-116, and Arg-122 (all TMD3) (shown in Fig. 3b). Similar to Seq1, Sim reveals a single line of hydrophilic residues due to hydrophilic residues of TMD3 (Fig. 3c).

Tetramer assembly without loops

Assembling T-Seq1 a structure with a minimum energy (−4710.72 kcal mol^{-1}, Table 2) is obtained for inter-monomer distance from the centers of mass of the monomers of 2.375 nm (Suppl. Fig. 2, I). The structure adopts a rotational angle of 320° and a tilt of 9°. TM2 is the pore lining with only one hydrophilic residue Tyr 91 (highlighted in Fig. 4a) facing the pore.

Alignment of T-Seq2 shows a Lennard-Jones type pattern for the low energy values with the lowest value −4724.84 kcal mol^{-1} of 2.15 nm interhelical distance (Suppl. Fig. 2, II). Although in T-Seq1 TMD2 is facing

Table 1 Lowest and second lowest energy structures of the dimer and finally the monomeric structures. Data represent the interhelical distance, rotational angle of each of the individual TMDs and the averaged (overall TMDs) tilt angle, as well as the energy calculated with MOE

Method	Distance (nm)	Angle 1 (°)	Angle 2 (°)	Angle 3 (°)	Tilt (°)	Energy (kcal/mol)
Seq1	1.0	40	340		8	−503.6
(TMD3 + TMD2)						
Seq1	1.20	260	220		−2	−878.1
((TMD3 + TMD2) + TMD1)	1.15	180	240		−10	−865.1
Seq2	0.95	340	300		0	−349.6
(TMD1 + TMD2)						
Seq2	1.175	0	40		2	−863.5
((TMD1 + TMD2) + TMD3)						
Sim	1.075	60	180	340	−4	−730.5
(TMD1 + TMD2 + TMD3)	1.075	40	100	80	4	−654.8
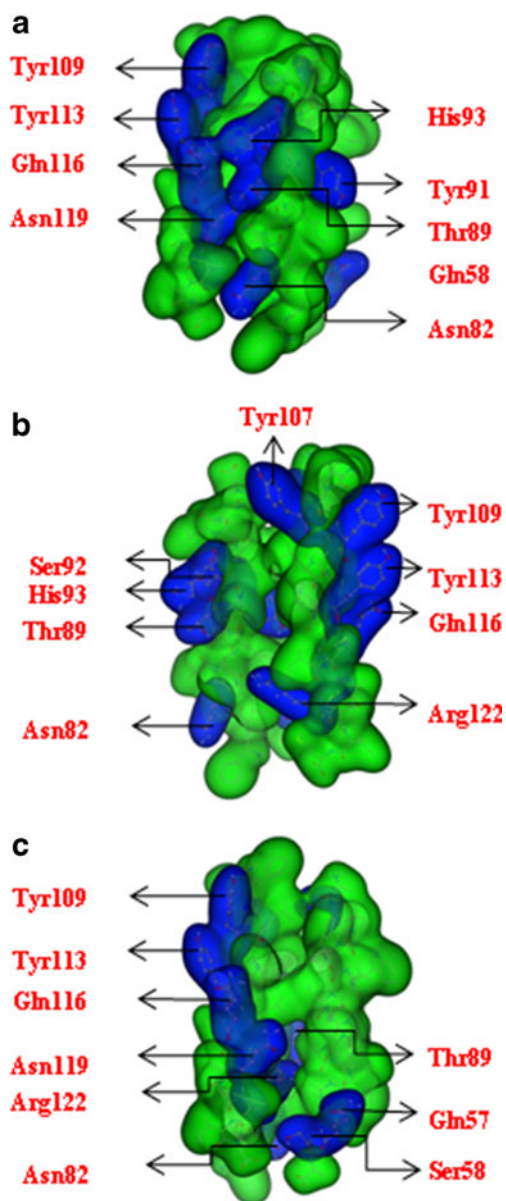
Fig. 3 Monomers according to the assembly protocol: Seq1 (**a**), Seq2 (**b**) and Sim (**c**). Hydrophilic residues are highlighted in blue, hydrophobic residues in green. All models are drawn in a 'Gaussian Contact' illustration (MOE)

the pore, similar to T-Seq2, the rotational angle is different in T-Seq2, leading to more hydrophilic residues inside the pore lumen (Asn-82, Thr-89, Ser-92, and His-93, Fig. 4b).

Alignment of T-Sim derives a lowest energy structure of −4294.2 kcal mol^{-1} with an inter monomer distance of 2.1 nm (Fig. 4c). A second low energy model (see Suppl. Fig. 2, III) does not expose any hydrophilic residues into the pore. In T-Sim TMD3 is pore lining, with several hydrophilic residues facing the pore (Tyr-109, Tyr-113, Gln-116, and Asn-119).

Tetramer assembly with added loops

In another approach the monomeric units are assembled in the presence of the loops between TMD1 and TMD2 as well as between TMD2 and TMD3. Assembling four copies of the Seq1-L monomer delivers a low energy structure with distances of around 2.025 nm (−5596.99 kcal mol^{-1}, T-Seq1-L) (Table 2). In T-Seq1-L bundle TMD2 is pore lining with Tyr-91 inside the pore lumen and His-93 facing outside the pore (Fig. 4d).

Assembling Seq2-L into a tetramer shows a low energy model with an monomer distance of 2.2 nm (−6136.76 kcal mol^{-1}), and a tilt angle of −36° T-Seq2-L is shown in Fig. 4e with two hydrophilic residues, Thr-89 and His-93, of TMD2 face the pore.

Screening the energy landscape of Sim-L, the model with the lowest energy (−5543 kcal mol^{-1}, T-Sim-L) has an inter monomer distance of 2.2 nm (data not shown). The tilt of its monomers adopts 21°. The lowest energy bundle, T-Sim-L, exposes hydrophilic residue Tyr 91 of TMD2 to the pore (Fig. 4f). Although the T-Sim-L is similar to T-Seq1-L with one with Tyr 91 inside the pore and His 93 outside the pore, the pore of T-Sim-L has more hydrophilic residues pointing into the pore than T-Seq1-L.

Comparing the energy values amongst the monomers reveals that Seq1 and Seq2 generate monomers with minimum energies around −860 kcal mol^{-1} to −880 kcal mol^{-1} whilst Sim generates monomers with higher values of around −650 kcal mol^{-1} and −730 kcal mol^{-1} (Suppl. Fig. 1). The bundle models reflect this trend independent of the presence of the loops (Suppl. Fig. 2). Whilst energies for bundles similar to T-Seq1 and T-Seq2 both are calculated to be around −4700 kcal mol^{-1}, the respective values for bundles similar to T-Seq1-L and T-Seq2-L show lower values for the bundles similar to T-Seq2-L: around −6100 kcal mol^{-1} (T-Seq2-L) versus −5600 kcal mol^{-1} (T-Seq1-L). The energy values for the bundles similar to T-Seq1-L are indistinguishable from those for bundles according to T-Sim-L. As a result, T-Seq2-L is the bundle with the low interaction energy.

MD simulations

All six tetrameric assembled structures of 3a from SARS CoV (Fig. 4) are run for 20 ns of a MD simulation embedded into a bilayer of POPC to equilibrate the structures further. The RMSD plot for Cα atoms of the tetrameric bundles without loops is shown in Fig. 5a. The data reveals a progressive rising for all structures and consequent stable fluctuation after the first 5 ns (Fig. 5a, I). All RMSD values remain in a range of 0.1 – 0.3 nm. In order to know how each TMD affects the stabilization of the structure, the RMSD of each TMD for the three bundles

Table 2 Lowest energy structures of the tetramer generated from monomers listed in Table 1. Data represent inter monomer distances, rotational and tilt angles, as well as the interaction energy calculated with MOE. The TMD of each of the monomer facing the pore is listed for each of the tetramers

Method	Distance (nm)	Angle (°)	Tilt (°)	Energy (kcal/mol)	Pore lining
T-Seq1	2.375	320	9	−4710.7	TMD2
T-Seq2	2..5	180	36	−4724.8	TMD2
T-Sim	2.10	190	24	−4294.2	TMD3
T-Seq1-L	2.025	310	24	−5596.9	TMD2
T-Seq2-L	2.20	150	−36	−6136.7	TMD2
T-Sim-L	2.20	305	21	−5543.0	TMD2

are shown individually. For T-Seq1 the RMSD of all TMDs are within the same range of 0.2 – 0.3 nm. (Fig. 5a, II). The RMSD values for TMD1 and TMD3 of T-Seq2 are higher (~ 0.24 nm for TMD1, and ~0.26 nm for TMD3) than for TMD2 (~ 0.15) (Fig. 5a, III). The same situation can also be found in T-Sim with TMD3 pore lining (RMSD ~ 0.19 nm) and TMD1 (~ 0.23 nm) and TMD2 (~ 0.25 nm) at

the outside of the bundle (Fig. 5a, IV). Super positioning the final structure (green, Fig. 6a-c) with the initial structures (red, Fig. 6a-c) indicates the result of the RMSD calculations in as much as the bundles do not deviate from each other very much, but show a pattern that the non-pore lining TMDs experience larger deviation from the initial structure than the pore lining residues.

Fig. 4 Tetramers according to the assembly protocols without loops: T-Seq1 (**a**), T-Seq2 (**b**), T-Sim (**c**); tetramers assembled with loops added after monomer assembly: T-Seq1-L (**d**), T-Seq2-L (**e**), T-Sim-L (**f**)
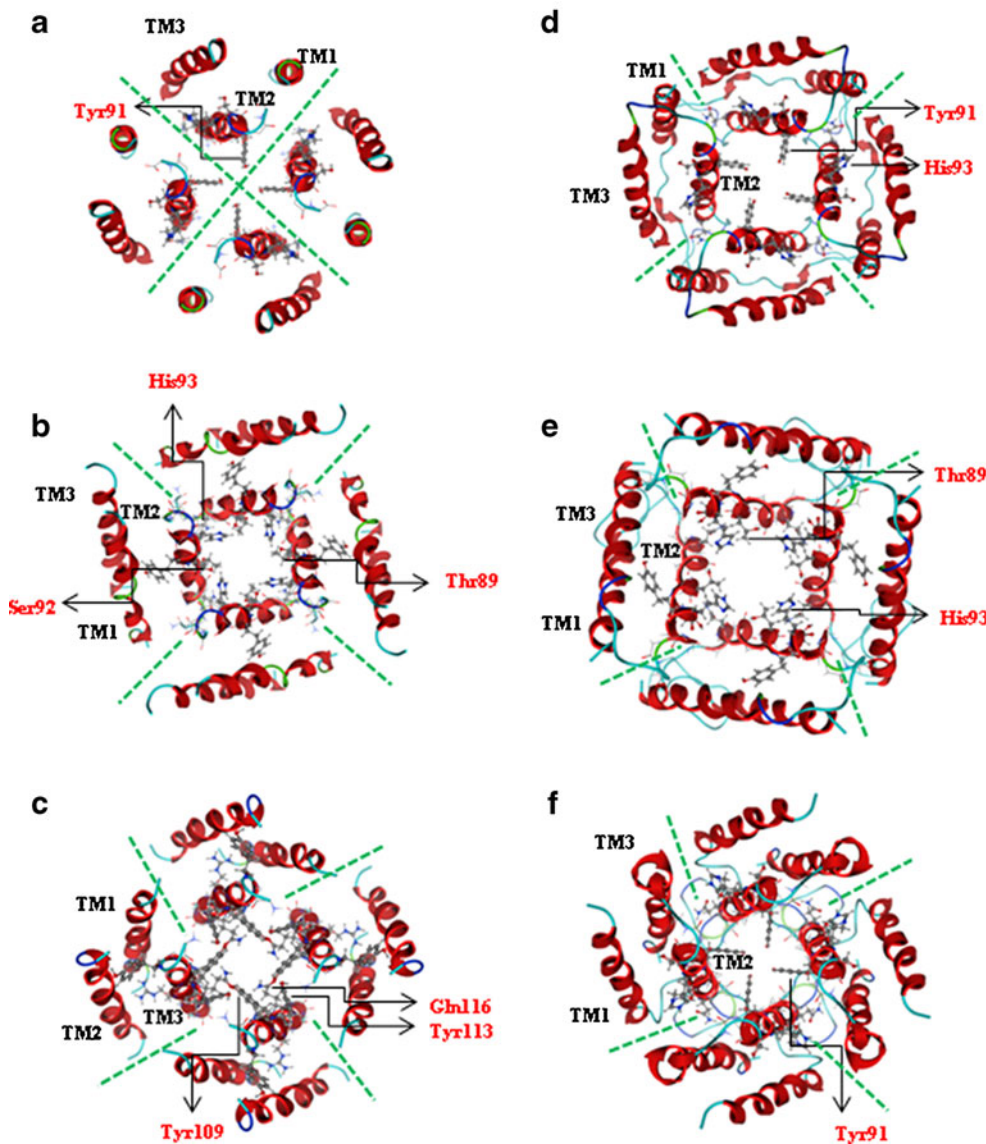
RMSD values for the bundles with loops indicate deviations in the range of 0.35 – 0.5 nm (Fig. 5b, I). T-Seq1; The large deviation is due to TMD1 in T-Seq1-L (~ 0.43 nm) and T-Seq2-L (~ 0.45 nm) shown in Fig. 5b, II and III. TMDs 2 and 3 in both bundles almost not deviate from each other. The RMSD values for TMDs of T-Sim-L are in a close range (0.24 ~ 0.30 nm) (Fig. 5b, IV). There is a tendency for increased values in the order TMD2 < TMD1 < TMD3. Indicating TMD2, which is pore lining to exhibit the lowest deviation. The superposition of the initial and final bundle for the structures with loops reflect the RMSD data that at least one of the TMD outside the pore has a large deviation, most likely TMD1. Less deviation is observed for the second outer TMD and the pore lining TMD.

Pore-radius analysis

The pore radii of the first 25 structures (covering five hundred pico second simulation in steps of 20 ps, Fig. 7, light lines) are compared to the radii derived toward the end of the simulation, taking the last 25 structures in steps of 20 ps for all the bundles (Fig. 7, thick lines). For T-Seq1 bundle, inside the membrane there are three local minima in the initial structure (Fig. 7a, thin line), caused by rings of Phe-87 (at −1.5 nm), Tyr-91 (at −0.5 nm), and Leu-94 (at 0.6 nm). The minimum pore radius is at Tyr-91, about 0.02 nm. Toward the end of the simulation only the region around Leu-94 is closed causing a minimum pore radius of 0.05 nm. For T-Seq2 (Fig. 7b), minima are caused by His-93 at position 0.3 nm and Leu-96 (at 1.2 nm). The minimum pore radius is at His-93, with about 0.04 nm. After 20 ns the pore radius is calculated to be around 0.02 nm around both, His-93 and Leu-96. T-Sim minima cover the stretch along Gln-116 (position −0.7 nm), Tyr-113 (position −0.1 nm), and Tyr-109 (position 1.0 nm) in the initial configuration (Fig. 7c). The minimum pore radius is at Gln-116, with about 0.026 nm. At the end of the simulation the entire stretch around Tyr-113 to Gln-116 retains a narrow pore passage with even Phe-105 at position 1.75 nm closing in at the mouth of the pore inducing almost a closure of the pore (minimum radius 0.03 nm).

The starting structure of T-Seq1-L bundle indicates a very narrow passage around Tyr-91 at position 0.45 nm with a minimum radius of 0.04 nm (Fig. 7d, thin line). After 20 ns the whole pore collapses and the minimum pore radius around Tyr-91 is at 0.013 nm (Fig. 7d, thick line). In T-Seq2-L a smallest pore radius is found around Leu-85 (position −1.1 nm) with about 0.15 nm (Fig. 7e). Two more space confinements are around Thr-89 (position 0.0 nm) with a radius of 0.2 nm and His-93 (position 0.85 nm) adopting a radius of 0.4 nm. During the simulation pore

Fig. 5 Root mean square deviation (RMSD) of the Cα backbones of ▶ the bundle structures referring to the starting structure. T-Seq1 (gray), T-Seq2 (light gray) and T-Sim (black in aI) are shown. The respective RMSD values for the individual TMDs of each simulation (TMD1 in light gray, TMD2 in gray, TMD3 in black) are shown separately (aII-IV). RMSD values of the bundles including the loops are shown for T-Seq1-L, TSeq2-L and T-Sim-L (**b**). Color coding and arrangement of the panels like in (**a**)

confines around Thr-89 at around −0.2 nm with a radius of 0.04 nm. For T-Sim-L the minimum pore radius of initial average structure is located at Tyr-91 at position 0.8 nm with a radius of 0.04 nm (Fig. 7f). At the end of the simulation the tyrosines have closed the pore. Constriction is at 1.0 nm due to the flexibility of the aromatic side chains.

Water molecules trajectories analysis

Water molecules do show three different kind of behaviors, (i) they get trapped in the pore found for T-Seq1 (data not shown), T-Sim-L, and T-Seq2-L (ii) they enter the pore on either side and escape on the same side found especially for T-Seq2 and T-Sim, and T-Seq1-L (iii) water molecules traverse the pore completely as found only for T-Seq2-L (5 water molecules in total). Adding the loop to the bundles results in pores with the likely hood of enabling a water passage across the bundle.

Discussion

Biological considerations

Experiments with 3a have identified the protein as a tetrameric unit enabling ion flux across the plasma membrane of infected Xenopus oocytes [4] which can also be inhibited by emodin [37]. Based on the experimental evidence the idea is to suggest a potential channel assembly based on experience in assembling smaller channel forming proteins [15, 31, 38–40]. Similar to other channel forming proteins such as Vpu from HIV-1 [41], also 3a is reported to interact with host factors [42]. Therefore these proteins are also called accessory proteins. The term implies that the presence of the protein helps the virus, but the virus is not dependent on it. Based on electrophysiological measurements the formation of channels cannot be ruled out at this stage and has to be considered also for drug development.

Considerations about the assembly protocol

A specific protocol is used to generate the tertiary structure of the TMD of a membrane protein [31]. It takes the secondary structural elements of the TMD which are helices
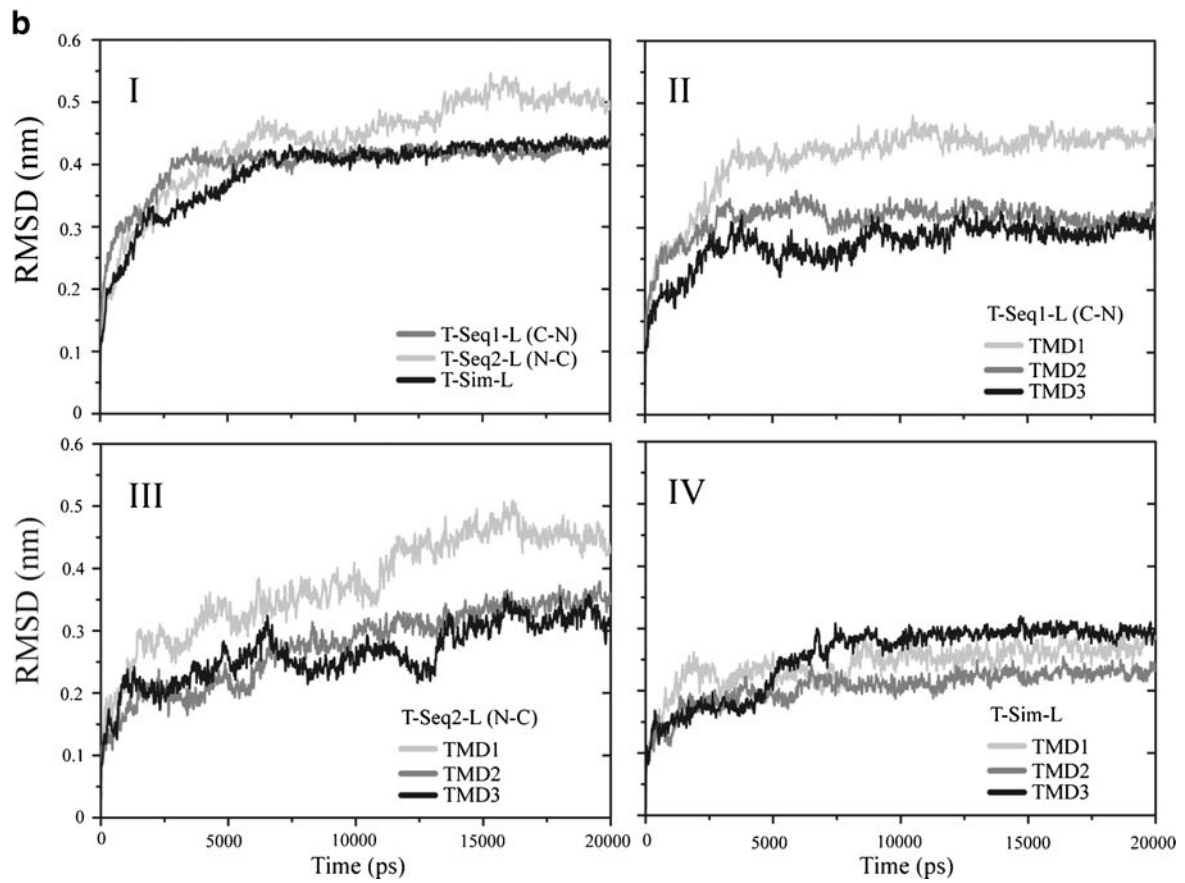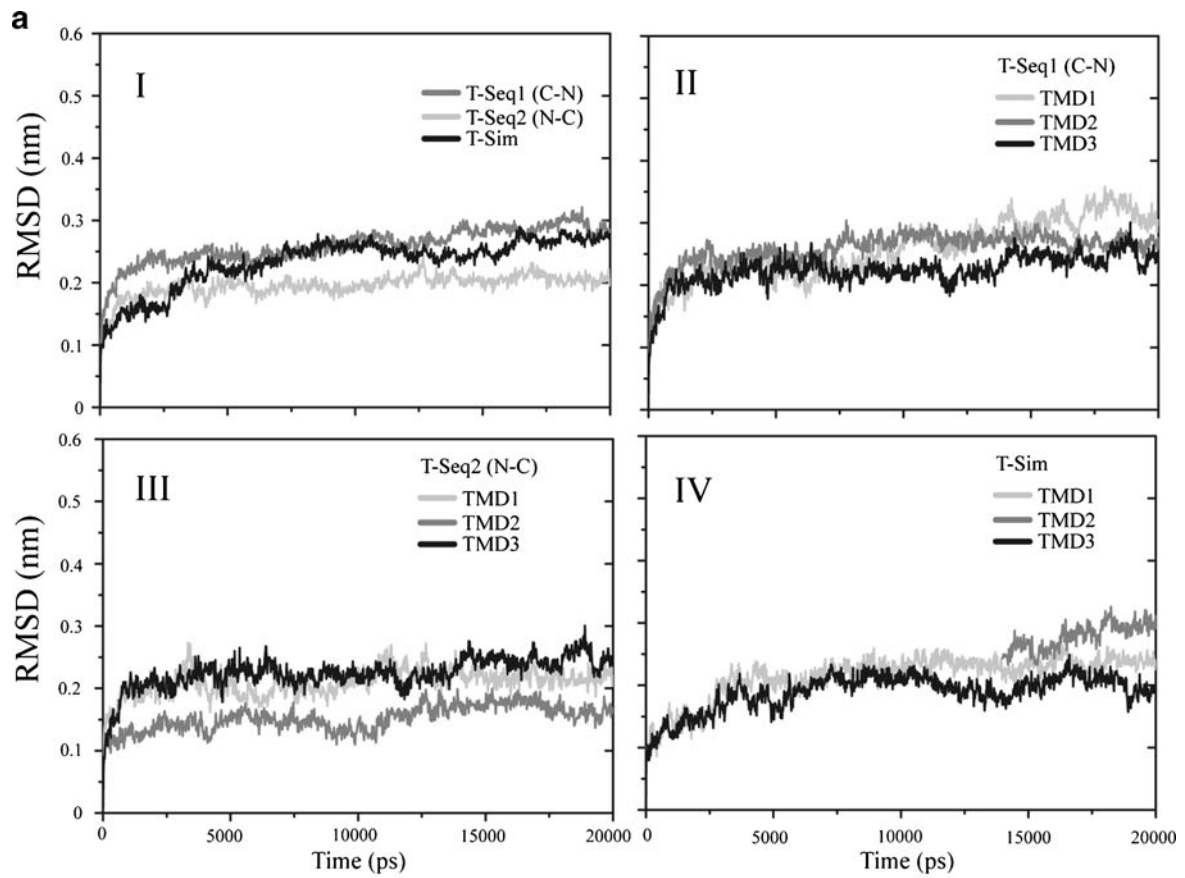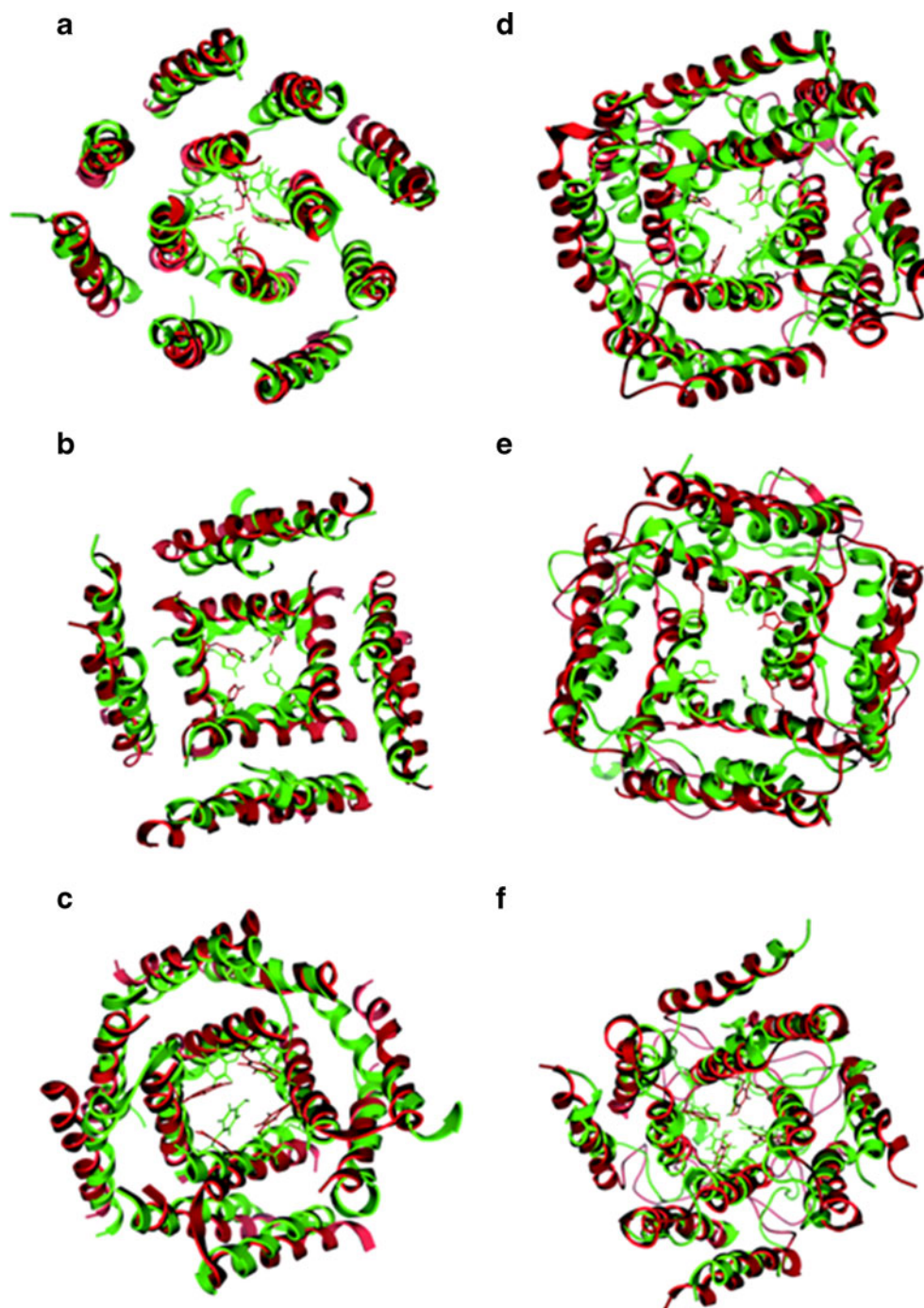
Fig. 6 Models of T-Seq1 (**a**), T-Seq2 (**b**) and T-Sim (**c**), T-Seq1-L (**d**), T-Seq2-L (**e**) and T-Sim-L (**f**) are shown in their starting conformation (green) and after 20 ns of MD simulation (red)


in this study and screens the interactions of these helices in 2D. Upon each positioning in 2D the potential orientation of the side chains at each position is taken from rotational library integrated in the program MOE. Each position is allowed to relax via energy minimization prior to energy calculations. Screening in 3D with a rigid body approach as done by other programs (e.g., [43]) has been omitted due to biological reasons as vertical movement of TMDs within a lipid bilayer is very much limited. It is anticipated that, e.g., adjustment to lipid dynamics is rather achieved by changes in tilt angles which is taken care of in the present assembly

approach. With the assembly protocol at hand it is possible to evaluate different kind of routes of assembly.

Assembly of membrane proteins and especially the TMDs can go two ways, either they are done 'ab initio', or they are done taking biological considerations into account. The first approach has been demonstrated to deliver results on other viral channel proteins which are in agreement with experiments [31]. Another approach is to assume biological pathways such as the TMDs once released from the translocon assembly step-by-step, in a sequential way. After another short period of time they find the other monomers to assemble finally in the

Fig. 7 Pore radii calculated using the software HOLE [56]. The values of the first 25 structures, covering 500 ps simulation in steps of 20 ps, are averaged and depicted in light lines. A similar average has been calculated covering the last 500 ps of the simulations (thick lines). Models of T-Seq1 (a), T-Seq2 (b) and T-Sim (c), T-Seq1-L (d), T-Seq2-L (e) and T-Sim-L (f) are shown
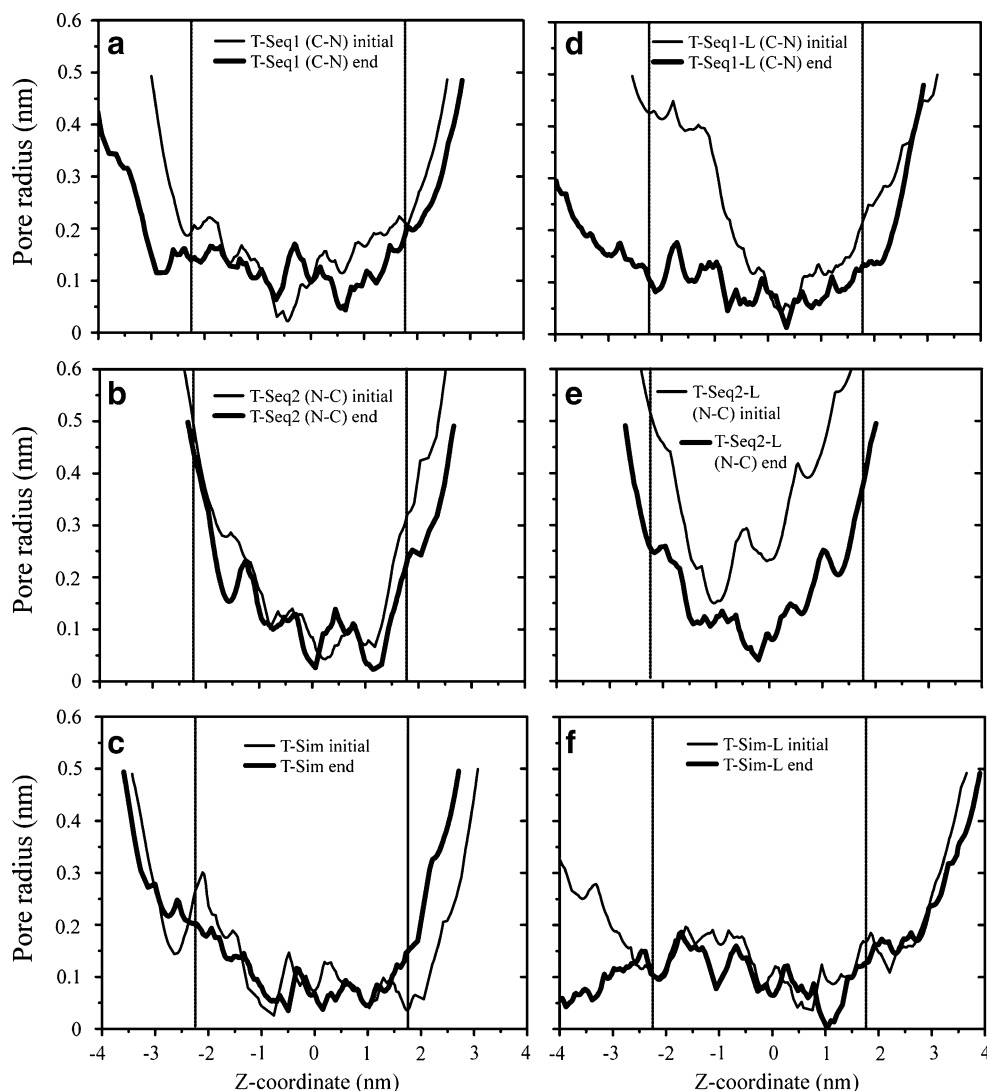
functional form. In the protocol described at the stage of assembling the tetramer, the concerted protocol is used. A sequential assembly at this stage, however, does not need to be ruled out. Assembly at this level may follow another biological pathway: The monomer can be in equilibrium between "free" and "raft" or "protein attached" states. Raft association has been proposed for M2 [44] and Vpu [45]. Thus also a raft attached state could be the seed for assembly of more monomers. In addition, the same scenario could be followed attaching to a host factor first or even to generate the covalent link between two of these monomers. All of these routes would be necessary to be taken into account. In the lack of any information about these scenarios the concerted assembly at this stage seems to be reasonable. It is assumed that the approach samples all low energy structures which inevitably impose constraints also on the "biological" pathways.

During the sequential assembly two routes are assumed, from the C to N termini and the opposite direction. The assembly route from C to N termini reflects the idea that TMD1 escapes the translocon first, 'diffuses' away and

allows the other two TMDs to be assembled first. This idea may be synonymous for a "loose" packing of the helices. The opposite route takes its rationale from the consideration that TMD1 may be retained near or at the translocon despite the longer loop between the TMD1 and the consecutive TMD2. Consequently TMD2 is manufactured and assembled with TMD1, followed by the assembly of TMD3. This route could be seen as a "constraint" packing.

Bundle and pore structure

All bundles in common are a pore lining TMD2 except for the bundle without loops built from the monomer using the simultaneous assembly protocol (T-Sim). TMD2 as the pore lining domain creates a Tyr-only (using Seq1 and Sim with loops) and a His/Tyr (Seq2) motif within the pore, whilst TMD3 creates a Tyr/Gln motif (using Sim without loops). A histidine within a pore has been found for M2 from influenza A [46, 47] and is proposed for p7 from HCV [39]. Tyrosines lining the pore may rather be unusual. Tyrosines

may catch a cation via cation/π interaction [48] and impose an ion trap along the pathway. Together with histidine this energy we think may be overcome making the bundle derived from Seq2 protocol the most likely one. Similar to M2 in respect to the number of monomers it seems to be likely that 3a may even be conducting protons rather than ions or should at least be pH dependent in its mechanism of function similar to the same proposal for p7 [39]. The Tyr/Gln motif may adopt the same mechanism as assumed for the bundle with the Tyr/His motif. At this stage it cannot be discriminated which motif would be the most effective one in respect to ion or proton conductance.

In a configuration of TMD2 the pore lining domain and TMD3 at the outside allows conformational freedom to enable covalent Cys-Cys linkage within the extramembrane part (Cys-133 in [4]) of two monomers without constraining the packing of the overall bundle in respect to the pore lining configuration.

Previously we have assembled in simultaneous mode, and got TMD3 pore lining. With a 'biological route' we suggest TMD2 pore lining.

Bundle dynamics

The results from the short equilibration dynamics of the bundles without loops deliver the picture that the inner helices of the bundles remain constrained relative to the TMDs at the outside of the bundle facing the lipid environment (T-Seq2, T-Sim). In all simulations of the bundles generated with the loops TMD1 shows the largest deviation from the starting structure whilst the values for TMD3 go almost in concord with those for TMD2 (T-Seq1-L, T-Seq2-L). This suggests that the assembly protocol delivers a structurally stable pore motif whilst the outer TMDs still need an extended equilibration. With the outer TMDs adjusting during the MD simulation, the pore lining TMDs are unaffected by the dynamics of the outer TMDs for the bundles without loop. It further implies that the short loop between TMD2 and TMD3 restrains the dynamics of TMD3. The findings suggest that the outer TMDs could be susceptible and allowing for some dynamics without affecting the inner helices.

In respect to the dynamics of the TMDs, analysis of the temperature (B) factor a series of crystal structures of known channel and pore proteins reveals a pattern in which helical TMDs surrounded by other helical TMDs show lower temperature (B) factors [49–53]. In the case of the mechanosensitive channel [54], the closed state model of pentameric ligand gated ion channel (LGIC) [49] and the glutamate receptor [51] a similar gradient of the temperature (B) factor for the TMDs across the membrane exists. For the mechanosensitive channel lower factors are found in the center of the TMDs and higher factors to both sides

whilst for pLGIC and the glutamate receptor the temperature factor decreases within the TMDs toward the extra-membrane domain of the channel. These data suggest that central TMDs adopt some rigidity whilst outer TMDs allow for some dynamics.

Water molecules in the pore

During the short equilibration the pore radius in all models fall below the radius of a sodium ion (e.g., 0.1 nm [55]) implying sever constraints onto the putative passage of ions. Only the bundle generated according to Seq2 with loops (T-Seq2-L) allows some water molecules to traverse. The water molecules remain on the level of the ring of His-93 for several ns before they leave the place in the other direction. All bundles have in common that not only hydrophilic stretches but also the rings of tyrosine impose special constrains on the passage through the pore. The findings for T-Seq2-L with water molecules crossing the pore and tyrosines restricting the pore it is likely that T-Seq2-L is the bundle of choice in this study. It may further underpin the suggestion of 3a to be proton conducting or at least sensitive to and triggered by the pH of the environment.

The lack of a continuous water column, which exists over the entire simulations in any of the bundles, imposes the question what are the necessities to generate and maintain such a column. At this stage it is speculated that ions are necessary to "stabilize" the pore and similar to the finding for the K^+-channel are essential for ion conductance.

At this stage any conductance of substrates has to be ruled out making the protein rather more ion channel like than pore like.

Role of the loops

Throughout the protocol we do not find a major impact of the loops on the structural modeling. The only exception is that in T-Sim TMD3 is suggested to be pore lining. However in the light of missing dynamics of the loops during assembly T-Sim may be rather a conformational exception. This underpins the idea that structural features can be independently modeled from the rest of the protein. Any extramembrane parts can be added after assembly. Possibly proteins are built in either of the environments, hydrophilic or hydrophobic, and then assembled. This leaves the question of the dynamics of the linker region between these two segments open for debate.

It is evident that the bundles with loops added have lower energy than those without loops. This is an indication that the addition of the loops improves the stability of the bundle.

Conclusions

Modeling of a membrane protein from ab initio conditions delivers a reasonable model of 3a prior to experimental calculations. Model generation is based on a combination of pure energetic considerations and the implementation of biological manufacturing praxis. As expected the computational approach delivers not a single result but the plurality can be reduced by considering further calculations on the proposed structural models. At the current level of calculations it is suggested that 3a adopts a bundle structure with TMD2 facing the putative pore albeit a TMD3 pore lining cannot be completely ruled out. The configuration delivers a Tyr and/or His motif to line the pore. It is further concluded based on the low pore radii generated by the protocol that ions embedded within the pore may be necessary to stabilize the pore and enabling ion flux. With histidine as part of the pore motif, 3a may also be a proton channel or at least sensitive or triggered by the pH around it. The pore architecture as presented would rule out 3a to be a substrate conducting pore.

Short equilibration runs using MD simulations are indicative for an excellent packing of the inner helices. The outer TMDs still need an extended equilibration to adjust for the bundle architecture.

With the more complex architecture 3a must be able to harbor a more precise activation mechanism. With this the role of the channel protein could be more specific and triggered by a more specific modulation mechanism underpinning is status as an ion/proton channel rather than a pore.

References

1. Marra MA, Jones SJM, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YSN, Khattra J, Asano JK, Barber SA, Chan SY, Cloutier A, Coughlin SM, Freeman D, Girn N, Griffith OL, Leach SR, Mayo M, McDonald H, Montgomery SB, Pandoh PK, Petrescu AS, Gordon Robertson A, Schein JE, Siddiqui A, Smailus DE, Stott JM, Yang GS, Plummer F, Andonov A, Artsob H, Bastien N, Bernard K, Booth TF, Bowness D, Czub M, Drebot M, Fernando L, Flick R, Garbutt M, Gray M, Grolla A, Jones S, Feldmann H, Meyers A, Kabani A, Li Y, Normand S, Stroher U, Tipples GA, Tyler S, Vogrig R, Ward D, Watson B, Brunham RC, Krajden M, Petric M, Skowronski DM, Upton C, Roper RL (2003) The genome sequence of the SARS-Associated Coronavirus. Science 300:1399–1404

2. Zeng R, Yang R-F, Shi M-D, Jiang M-R, Xie Y-H, Ruan H-Q, Jiang X-S, Shi L, Zhou H, Zhang L, Wu XD, Lin Y, Ji YY, Xiong L, Jin Y, Dai EH, Wang XY, Si BY, Wang J, Wang HX, Wang CE, Gan YH, Li YC, Cao JT, Zuo JP, Shan SF, Xie E, Chen SH, Jiang ZQ, Zhang X, Wang Y, Pei G, Sun B, Wu JR (2004) Characterization of the 3a protein of SARS-associated coronavirus in infected Vero E6 cells and SARS patients. J Mol Biol 341:271–279

3. Tan YJ, Teng E, Shen S, Tan THP, Goh PY, Fielding BC, Ooi EE, Tan HC, Lim SG, Hong W (2004) A novel severe acute respiratory syndrome coronavirus protein, U274, is transported to the cell surface and undergoes endocytosis. J Virol 78:6723–6734

4. Lu W, Zheng BJ, Xu K, Schwarz W, Du L, Wong CKL, Chen J, Duan S, Deubel V, Sun B (2006) Severe acute respiratory syndrome-associated coronavirus 3a protein forms an ion channel and modulates virus release. Proc Natl Acad Sci USA 103:12540–12545

5. Chen CY, Ping YH, Lee HC, Chen KH, Lee YM, Chan YJ, Lien TC, Jap TS, Lin CH, Kao LS, Chen YMA (2007) Open reading frame 8a of the Human severe acute respiratory syndrome coronavirus not only promotes viral replication but also induces apoptosis. J Infect Dis 196:405–415

6. Fischer WB, Sansom MSP (2002) Viral ion channels: structure and function. Biochim Biophys Acta 1561:27–45

7. Fischer WB, Hsu HJ (2011) Viral channel forming proteins - modelling the target. Biochim Biophys Acta 1808:561–571

8. Fischer WB, Krüger J (2009) Viral channel forming proteins. Int Rev Cell Mol Biol 275:35–63

9. Allen H, McCauley J, Waterfield M, Gethering M (1980) Influenza virus RNA segment 7 has the coding capacity for two polypeptides. Virology 107:548–551

10. Holsinger LJ, Nichani D, Pinto LH, Lamb RA (1994) Influenza A virus M2 ion channel protein: a structure-function analysis. J Virol 68:1551–1563

11. Mould JA, Li HC, Dudlak CS, Lear JD, Pekosz A, Lamb RA, Pinto LH (2000) Mechanism for proton conduction of the M_2 ion channel of influenza A virus. J Biol Chem 275:8592–8599

12. Lin T, Schroeder C (2001) Definitive assignment of proton selectivity and attoampere unitary current to the M2 ion channel protein of influenza A virus. J Virol 75:3647–3656

13. Schubert U, Ferrer-Montiel AV, Oblatt-Montal M, Henklein P, Strebel K, Montal M (1996) Identification of an ion channel activity of the Vpu transmembrane domain and its involvement in the regulation of virus release from HIV-1-infected cells. FEBS Lett 398:12–18

14. Ewart GD, Sutherland T, Gage PW, Cox GB (1996) The Vpu protein of human immunodeficiency virus type 1 forms cation-selective ion channels. J Virol 70:7108–7115

15. Chen CC, Krüger J, Sramala I, Hsu HJ, Henklein P, Chen YMA, Fischer WB (2011) ORF 8a of severe acute respiratory syndrome coronavirus forms an ion channel: experiments and molecular dynamics simulations. Biochim Biophys Acta 1808:572–579

16. Griffin SDC, Beales LP, Clarke DS, Worsfold O, Evans SD, Jäger J, Harris MPG, Rowlands DJ (2003) The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, amantadine. FEBS Lett 535:34–38

17. Pavlovic D, Neville DCA, Argaud O, Blumberg B, Dwek RA, Fischer WB, Zitzmann N (2003) The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. Proc Natl Acad Sci USA 100:6104–6108

18. Agirre A, Barco A, Carrasco L, Nieva JL (2002) Viroporin-mediated membrane permeabilization. Pore formation by nano-structural poliovirus 2B protein. J Biol Chem 277:40434–40441

19. de Jong AS, Wessels E, Dijkman HBPM, Galama JMD, Melchers WJG, Willems PHGM, van Kuppeveld FJM (2003) Determinants for membrane association and permeabilization of the coxsack-

ievirus 2B protein and the identification of the Golgi complex as the target organelle. J Biol Chem 278:1012–1021

20. Pervushin K, Tan E, Parthasarathy K, Lin X, Jiang F-L, Yu D, Vararattanavech A, Soong TW, Liu D-X, Torres J (2009) Structure and inhibition of the SARS Coronavirus envelope protein ion channel. PLoS Pathog 5:e1000511

21. Cook GA, Zhang H, Park SH, Wang Y, Opella SJ (2011) Comparative NMR studies demonstrate profound differences between two viroporins: p7 of HCV and Vpu of HIV-1. Biochim Biophys Acta 1808:554–560

22. Popot JL, Engelman DM (1990) Membrane protein folding and oligomerization: the two-stage model. Biochemistry 29:4031–4037

23. Engelman DM, Chen Y, Chin CN, Curran AR, Dixon AM, Dupuy AD, Lee AS, Lehnert U, Matthews EE, Reshetnyak YK, Senes A, Popot JL (2003) Membrane protein folding: beyond the two stage model. FEBS Lett 555:122–125

24. White SH, Wimley WC (1999) Membrane protein folding and stability: physical principles. Annu Rev Biophys Biomol Struct 28:319–365

25. Sansom MSP, Kerr ID (1993) Influenza virus M_2 protein: a molecular modelling study on the ion channel. Protein Eng 6:65–74

26. Kukol A, Adams PD, Rice LM, Brunger AT, Arkin IT (1999) Experimentally based orientational refinement of membrane protein models: a structure for the influenza A M2 H^+ channel. J Mol Biol 286:951–962

27. Cordes F, Kukol A, Forrest LR, Arkin IT, Sansom MSP, Fischer WB (2001) The structure of the HIV-1 Vpu ion channel: modelling and simulation studies. Biochim Biophys Acta 1512:291–298

28. Bu L, Im W, Brooks CL III (2007) Membrane assembly of simple helix homo-oligomers studied *via* molecular dynamics simulations. Biophys J 92:854–863

29. Psachoulia E, Marshall DP, Sansom MSP (2010) Molecular dynamics simulations of the dimerization of transmembrane α-helices. Acc Chem Res 43:388–396

30. Bowie JU (2005) Solving the membrane protein folding problem. Nature 438:581–589

31. Krüger J, Fischer WB (2009) Assembly of viral membrane proteins. J Chem Theory Comput 5:2503–2513

32. Wahba K, Schwab D, Bruinsma R (2010) Statistical mechanics of integral membrane protein assembly. Biophys J 99:2217–2224

33. Chandrasekhar I, Kastenholz M, Lins RD, Oostenbrink C, Schuler LD, van Gunsteren WF (2003) A consistent potential energy parameter set for lipids: dipalmitoylphosphatidylcholine as a benchmark of the GROMOS96 45A3 force field. Eur Biophys J 32:67–77

34. Krüger J, Fischer WB (2008) Exploring the conformational space of Vpu from HIV-1: a versatile and adaptable protein. J Comput Chem 29:2416–2424

35. Xiang Z, Soto CS, Honig B (2002) Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. Proc Natl Acad Sci USA 99:7432–7437

36. Soto CS, Fasnacht M, Zhu J, Forrest L, Honig B (2008) Loop modeling: sampling, filtering, and scoring. Proteins 70:834–843

37. Schwarz S, Wang K, Yu W et al. (2011) Emodin inhibits current through SARS-associated coronavirus 3a protein. Antiviral Res (ahead of print)

38. Cordes FS, Tustian AD, Sansom MS, Watts A, Fischer WB (2002) Bundles consisting of extended transmembrane segments of Vpu from HIV-1: computer simulations and conductance measurements. Biochemistry 41:7359–7365

39. Patargias G, Zitzmann N, Dwek R, Fischer WB (2006) Protein-protein interactions: modeling the hepatitis C virus ion channel p7. J Med Chem 49:648–655

40. Patargias G, Barke T, Watts A, Fischer WB (2009) Model generation of viral channel forming 2B protein bundles from polio and coxsackie viruses. Mol Membr Biol 26:309–320

41. Skasko M, Tokarev A, Chen C-C et al. (2011) BST-2 is rapidly down-regulated from the cell surface by the HIV-1 protein Vpu: evidence for a post-ER mechanism of Vpu-action. Virology (ahead of print)

42. Narayanan K, Huang C, Makino S (2008) SARS coronavirus accessory proteins. Virus Res 133:113–121

43. Ausiello G, Cesareni G, Helmer-Citterich M (1997) ESCHER: a new docking procedure applied to the reconstruction of protein tertiary structure. Proteins 28:556–567

44. Schroeder C, Heider H, Moncke-Buchner E, Lin TI (2005) The influenza virus ion channel and maturation cofactor M2 is a cholesterol-binding protein. Eur Biophys J 34:52–66

45. Ruiz A, Hill MS, Schmitt K et al. (2010) Membrane raft association of the Vpu protein of human immunodeficiency virus type 1 correlates with enhanced virus release. Virology (in press)

46. Schnell JR, Chou JJ (2008) Structure and mechanism of the M2 proton channel of influenza A virus. Nature 451:591–595

47. Cady SD, Schmidt-Rohr K, Wang J, Soto CS, DeGrado WF, Hong M (2010) Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. Nature 463:689–692

48. Dougherty DA (1996) Cation - π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. Science 271:163–168

49. Hilf RJC, Dutzler R (2008) X-ray structure of a prokaryotic pentameric ligand-gated ion channel. Nature 452:375–379

50. Hilf RJC, Dutzler R (2009) Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. Nature 457:115–119

51. Sobolevsky AI, Rosconi MP, Gouaux E (2009) X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. Nature 462:745–756

52. Mueller M, Grauschopf U, Maier T, Glockshuber R, Ban N (2009) The structrure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. Nature 459:726–730

53. Waight AB, Love J, Wang DN (2010) Structure and mechanism of a pentameric formate channel. Nat Struct Biol 17:31–37

54. Steinbacher S, Bass R, Strop P, Rees DC (2007) Structures of the prokaryotic mechanosensitive channels MscL and MscS. Curr Top Membr 58:1–24

55. Stein WD (1990) Channels, carriers and pumps. An introduction to membrane transport. Academic Press Inc, San Diego, CA

56. Smart OS, Neduvelil JG, Wang X, Wallace BA, Sansom MSP (1996) Hole: a program for the analysis of the pore dimensions of ion channel structural models. J Mol Graph 14:354–360

ORIGINAL PAPER

A Curtin–Hammett mechanism for the copolymerization of ethylene and methyl acrylate monomer using a PymNox nickel catalyst as revealed by DFT computational studies

Javier Ramos · Sonia Martínez · Víctor L. Cruz · Javier Martínez-Salazar

Abstract In this work, the copolymerization of ethylene and methyl acrylate (MA) as catalyzed by a new Ni-based PymNox organometallic compound was studied computationally. We recently tested the behavior of this type of catalyst in ethylene homopolymerization. Experimental results show that the unsubstituted catalyst Ni2 (aldimino PymNox catalyst) is unable to incorporate the MA monomer, whereas methyl-substituted Ni1 (acetaldimino PymNox catalyst) is able to achieve copolymerization. The reactivities of both catalysts were examined using density functional theory (DFT) models. Based on energy profiles calculated at the BP86 level, a Curtin–Hammett mechanism was proposed to explain the different reactivities of the catalysts in ethylene/MA copolymerization. Our results indicate that the methyl substituent Ni1 introduces additional steric hindrance that results in a catalyst conformation that is better suited to polar monomer incorporation. This model provides insights into the design of new catalysts to produce polar functionalized copolymers based on ethylene.

J. Ramos (✉) · V. L. Cruz · J. Martínez-Salazar
BIOPHYM, Departamento de Física Macromolecular,
Instituto de Estructura de la Materia, CSIC,
Serrano 113bis,
28006, Madrid, Spain
e-mail: j.ramos@iem.cfmac.csic.es

S. Martínez
Área de Informática Científica,
Secretaria General Adjunta de Informática (SGAI), CSIC,
Pinar 19,
28006, Madrid, Spain

Introduction

The copolymerization of a polar monomer with ethylene to give a functionalized polymer significantly modifies the properties of the ethylene homopolymer in terms of compatibility, adhesion and rheology [1]. Hence, the efforts of scientists working in the polyolefin field are currently focused on finding new synthetic routes for functionalized polymers under mild conditions.

Several copolymerization studies have examined the production of copolymers from ethylene and functionalized monomers using traditional Ziegler–Natta and metallocene catalysts [2–9]. However, as a consequence of the high oxophilicity of early transition metals, the activities of these catalysts are substantially reduced. Metal centers are deactivated by comonomers containing a heteroatom-bearing functional group (hydroxyl, ester, carboxy and so on), thus hindering the growth of the polymer chain.

Late transition metal catalysts are less oxophilic and are, at first glance, more tolerant of polar groups. However, late transition metal catalysts are best known to oligomerize ethylene and propylene [10–12], since these metals generally favor β-hydride elimination reactions over insertion ones. On the other hand, late transition metal catalysts have attracted more attention since the discovery by Brookhart and coworkers [13] of a new group of Ni(II) and Pd(II) catalysts bearing bulky aryl α-diimine ligands, which allow very efficient olefin polymerization. In these compounds, bulky ancillary ligands slow down chain-transfer reactions, leading to high molecular weight polymers. However, while

the most expensive palladium α-diimine catalysts are able to produce ethylene-acrylate or propene-acrylate copolymers under certain experimental conditions, their nickel α-diimine counterparts are easily deactivated by polar comonomers [14, 15].

In contrast, nickel catalysts based on neutral ligands exhibiting a mixed donor atom set (N,O or P,O) are good candidates for copolymerization catalysts due to their less electrophilic centers [16–19]. If we decrease the negative charge density at the metal center, we would expect to see an increase in activity, while the capacity of the above-mentioned catalysts to incorporate polar comonomers would be retained.

In line with this idea, we modified Grubbs' salicylaldiminate catalyst design by replacing the anionic aryloxide ligand with an electrically neutral but isostructural pyridine-N-oxide fragment. The resulting 2-iminopyridine-N-oxide ligand (PymNox) is isostructural to salicylaldiminate, but gives rise to a cationic polymerization system [20].

Scheme 1 shows the skeleton structure of the PymNox catalysts that have recently been synthesized and tested for the copolymerization of ethylene with methyl acrylate. Experimental results show that the activities of most complexes decrease when methyl acrylate comonomer is added to the reactor. The exception is the catalyst Ni1, which remains active a long time after the introduction of the comonomer to produce ethylene-methyl acrylate copolymer. NMR analysis suggests that acrylate units and short CH_2CO_2Me branches are inserted into the resultant copolymer [20].

So far, our approach, which combines experimental and simulation studies, has proven useful for understanding the mechanisms involved in polymerization processes [21–26]. Theoretical studies conducted by us and by other groups have clarified the polymerization mechanisms of ethylene [27–31], propylene [29–32], and the copolymerization of ethylene and polar comononers containing oxygen and nitrogen as heteroatoms [33–36] using nickel and palladium α-diimine and salicylaldiminate complexes.

Based on DFT studies, Michalak et al. [33, 36] showed that Ni(II) α-diimine catalysts are inactive in the copolymerization of ethylene and oxygen-containing polar monomers, such as methyl acrylate (MA) and vinyl acetate (VA), while palladium α-diimine complexes can catalyze copolymerization reactions. Energy profiles indicate that the polar monomers are bound by a carbonyl oxygen atom to
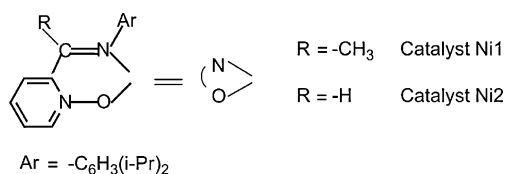
the nickel complexes, thus poisoning the active center. However, vinyl π-complexes are preferred for palladium catalysts, allowing comonomer insertion. It has been proposed that the origin of this difference is mainly steric. The same authors have shown that vinyl π-complexes are preferred over oxygen complexes for both metals in neutral salicylaldiminate ligands (Grubbs' catalysts). In addition, steric bulk ligands are necessary to weaken chelate bonds between the metal and the polar group that might otherwise poison the catalyst [36].

Here, we present a theoretical study of a new family of cationic 2-iminopyridine-N-oxide nickel (PymNox) complexes used in ethylene and MA copolymerization. Our study was designed to gain insight into the factors that control the mechanisms that allow ethylene and MA monomers to insert into the metal center, and to find a suitable explanation for the differences detected between two different PymNox complexes. These complexes can effectively be used as representative models of active and inactive catalysts in ethylene–MA copolymerization, respectively. The present work complements the results already reported by our group on ethylene homopolymerization using the PymNox system [37].

The article is organized as follows. In the next section, we introduce the computational methods and the nomenclature used in this paper. After this, we present the most relevant results concerning the geometries and energies of the copolymerization mechanism for both catalysts. We studied MA insertion, chelate stability, and isomerization reactions. In the "Discussion" section, a global analysis of the proposed Curtin–Hammett mechanism is performed. Finally, a summary of the main conclusions of this study is given.

Simulation details

Computational methods

All energies and geometries were calculated using the BP86 DFT method implemented in the Amsterdam Density Functional (ADF) package [38]. This method uses local-density approximations from Vosko–Wilk–Nusair [39], along with nonlocal corrections for energy exchange from Perdew [40] and correlation from Becke [41].

For all atoms, the innermost atomic shells were treated within the frozen core approximation to minimize the computational effort. The outermost shells were described using a triple-ζ basis set plus a polarization function on the nickel atom. A double-ζ basis set, augmented with a polarization function, was used for the remaining atoms.

Transition-state geometries were obtained by employing the LST (linear synchronous transit) method to locate an

Scheme 1 Ligand structure of 2-iminopyridine-N-oxide (PymNox)

estimated saddle point along the path from reactant to product. Subsequently, this "guessed" saddle point was completely optimized according to the negative or smaller eigenvector. Frequency calculations were performed to check the nature of the identified stationary points. The stationary points revealed a lack of imaginary frequencies, confirming the presence of true minima on the potential energy surface. Transition states were characterized by exactly one imaginary frequency, visualizing the corresponding eigenvector.

In the present model, the cocatalyst MAO creates the active site by alkylating the catalyst precursor and removing one of the methyl groups. The cationic nickel species, which has a vacant site in the coordination sphere, is then taken as the starting point for the polymerization reactions with the monomers. The growing alkyl chain was represented by an *n*-propyl group in the calculations.

Nomenclature

Scheme 1 shows the schematic cationic structures of the two catalysts studied, which are called Ni1 and Ni2, together with the schematic structure of the 2-iminopyridine-N-oxide ligand.

The different structures presented in this study are numbered starting from **1**. **1** corresponds to β-agostic *n*-propyl complexes, which are considered to be the resting states for the propagation chain reaction.

Due to the asymmetry of the PymNox ligand, two different paths must be considered in both catalysts, so there are two different vacant sites for the coordination and insertion processes (Scheme 2). The structures are *cis* isomers when the *n*-propyl chain is on the same side as the oxygen atom in the PymNox ligand. On the other hand, when the *n*-propyl is opposite to the O atom, the structures are the *trans* isomers.

In this sense, we also include a letter "c" for the structures created during the polymerization process that come from the initial *cis* isomers, irrespective of the alkyl chain position. Consequently, the letter "t" is used for those that come from the *trans* structures, as shown in Scheme 2.

On the other hand, during the copolymerization process, different paths are initiated depending on the monomer type and its coordination mode, as can be seen in Scheme 3. There are two monomers that compete for the active sites of the catalysts. The ethylene monomer can only coordinate in one way to form a π-complex. However, the methyl

acrylate monomer is able to bind through its vinyl group, forming an MA π-complex, and through its carbonyl O atom, producing an O-bound complex.

Furthermore, there are two possible vinyl coordination modes for the MA monomer. These are 1,2 or 2,1, depending on whether the secondary atom binds to the metal atom or to the Cα of the growing polymer chain, and these lead to the 1,2 π-complex or to the 2,1 π-complex, respectively.

Results

Previous results: ethylene polymerization

At this point, it is useful to summarize the main conclusions of our previous work on ethylene homopolymerization with these catalysts [37].

From a geometric point of view, we found some differences in the structures of the β-agostic resting states associated with both catalysts due to the steric hindrance introduced by the methyl substituent in the Ni1 catalyst [37].

Concerning the stability of these species, the *cis*-β-agostic alkyl conformers (**1c** in Scheme 5), were 1.5 and 2.5 kcal mol^{-1} less stable than the *trans* ones (**1t** in Scheme 4) for the Ni1 and Ni2 catalysts, respectively. However, ethylene homopolymerization insertion barriers were considerably lower for the *cis* pathway than for the *trans* one by about 6 kcal mol^{-1} for both catalysts. Due to this large difference between the insertion barriers, two *trans–cis* isomerization mechanisms were considered.

First, *trans–cis* isomerization between the β-agostic resting states was found to be very unlikely (it had a barrier >25 kcal mol^{-1}). On the other hand, isomerization of the ethylene π-complexes presented energy barriers that were considerably lower than the ethylene insertion barrier (4.8 vs. 15.7 for Ni1 and 8.0 vs. 12.2 kcal mol^{-1} for Ni2), so a Curtin–Hammett mechanism was proposed in this case. Therefore, the *trans*-π-complexes are isomerized to the *cis* one, and the insertion always takes place through this pathway.

Methyl acrylate (MA) complexation and insertion

We studied the coordination and insertion of MA monomer into the β-agostic alkyl resting state. As mentioned before, there are three different coordination modes for the MA monomer into the active species (see Scheme 3). Our calculations show that 2,1 insertions lead to more stable complexes than 1,2 insertions. Similar results were reported by Michalak et al. [37] for MA insertion using α-diimine nickel complexes. Thus, from now on, we only consider 2,1 insertion.
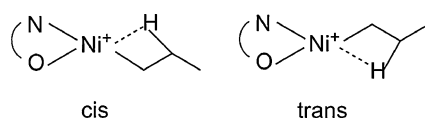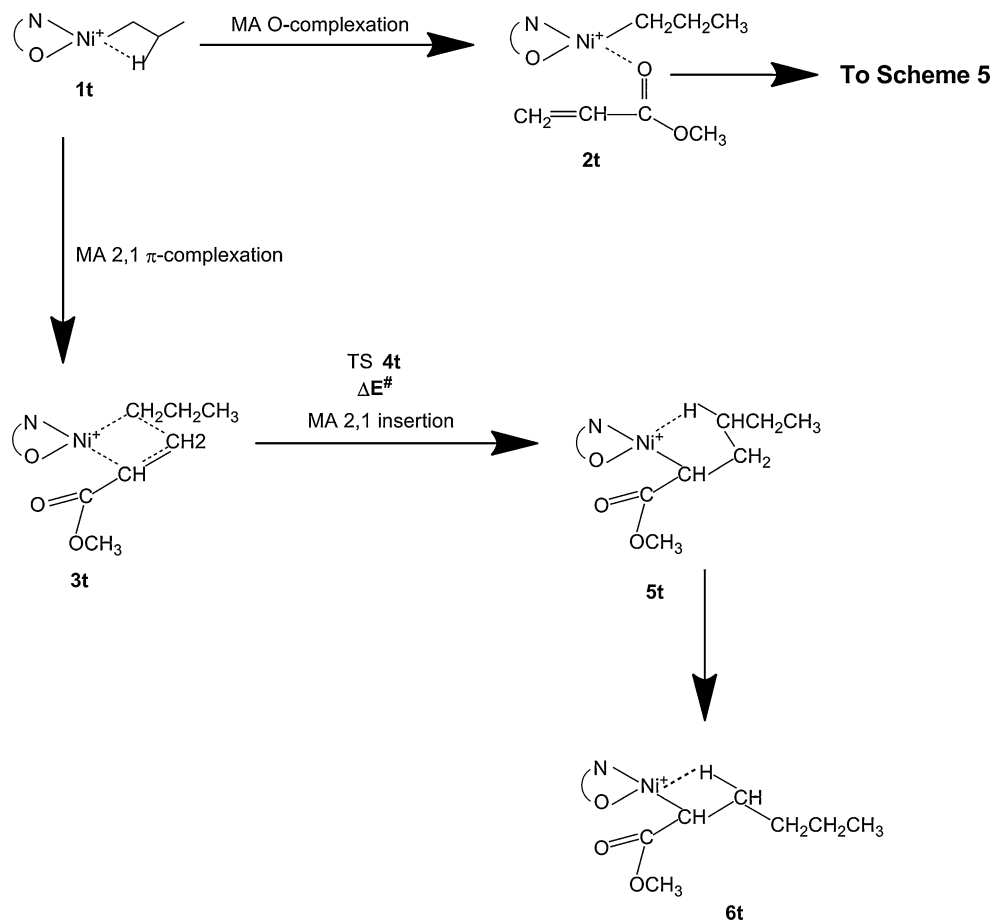

Scheme 2 The *cis* and *trans* isomers of the PymNox nickel catalysts
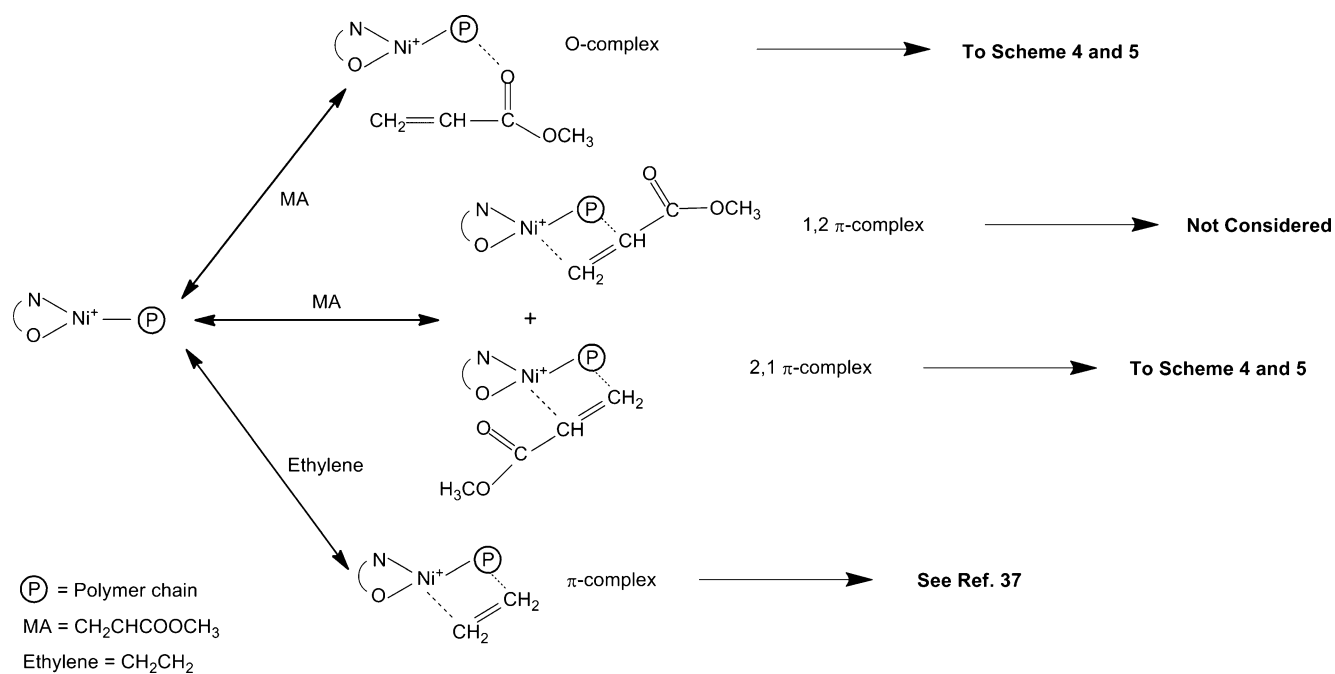
Scheme 3 Different coordination modes for ethylene and MA monomers into PymNox nickel catalysts

Scheme 4 Mechanism for MA polymerization via the *trans* pathway, as catalyzed by PymNox nickel catalysts

As was concluded in our previous work, a Curtin–Hammett mechanism is proposed for the ethylene insertion [37]. The *trans–cis* interconversion of the π-vinyl alkyl complexes occurs more rapidly than the monomer insertion step. In addition, the monomer insertion barrier is lower for the *cis* pathway. Bearing in mind these results, the *cis* pathway is always preferred. Thus, the **1t** geometries (Scheme 4) are the only isomers that can utilize the MA or ethylene monomers in the (co)polymerization process.

The O-bound coordination mode of the MA monomer into *trans*-alkyl β-agostic products leads to an O-complex (see the **1t → 2t** step in Scheme 4). The binding energies are −12.6 and −13.8 kcal mol^{-1} for Ni1 and Ni2 catalysts, respectively (see Table 1). On the other hand, the corresponding binding energies for the 2,1 π-vinyl coordination mode are −7.1 and −9.0 kcal mol^{-1} for Ni1 and Ni2 complexes, respectively (see the **1t → 3t** step in Scheme 4). In these π-vinyl complexes, the π-bond of the MA lies

perpendicular to the Ni–Cα bond. In summary, O-coordination is more favorable than 2,1-π-vinyl coordination by 5.5 and 4.8 kcal mol^{-1} for the Ni1 and Ni2 catalysts, respectively.

MA insertion from **3t** intermediates occurs through the transition states **4t**, which are 24.6 and 20.6 kcal mol^{-1} less stable than the corresponding π-vinyl complexes (**3t**) for both Ni1 and Ni2 catalysts, respectively. The direct insertion γ-agostic products (**5t** in Scheme 4) are −6.3 and −3.1 kcal mol^{-1} more stable than their corresponding π-complexes, (**3t** in Scheme 4). These evolve to the most stable β-agostic products (**6t** in Scheme 4), which are −10.6 and −11.0 kcal mol^{-1} more stable than the γ-agostic ones for both catalysts.

Isomerization processes

The γ-agostic and β-agostic products formed after MA insertion via the *trans* path are more stable than the reactants. However, the energy barrier for this insertion reaction is high for both catalysts (24.6 and 20.6 kcal mol^{-1} for Ni1 and Ni2, respectively). Furthermore, the O-bound coordination mode is favored over the 2,1-π-vinyl one. Thus, MA insertion seems to be very unlikely via this pathway for both catalysts. In fact, it appears to deactivate

Table 1 Energy differences in kcal mol^{-1} associated with ethylene/MA copolymerization processes catalyzed by PymNox complexes (see Schemes 4 and 5)

trans pathway (Scheme 4)		
Step	Ni1	Ni2
	ΔE	
1t + MA → 2t	−12.6	−13.8
1t + MA → 3t	−7.1	−9.0
3t → [4t]$^{\#}$	+24.7	+20.6
3t → 5t	−6.3	−3.1
5t → 6t	−10.6	−11.0
cis pathway (Scheme 5)		
Step	Ni1	Ni2
	ΔE	
1c + MA → 2c	−9.7	−13.2
1c + MA → 3c	−11.2	−10.6
2t → [2t–3c]$^{\#}$	<7.0[a]	<7.0[a]
2t→3c	+4.4	+3.1
2c → [1c + MA]$^{\#}$	+9.7	+13.2
2c → 3c	−1.5	+2.6
3c → [4c]$^{\#}$	+14.1	+10.5
3c → 5c	−2.0	−5.6
5c → 6c	−8.6	−7.2
6c → 8c	−15.7	−16.4
8c → [9c]$^{\#}$	n.a.[b]	+10.6
8c → 10c	−16.1	−18.2
10c → 11c	−9.3	−5.2
11c → 13c	−5.3	−8.1

Energy differences for the step A → B were calculated as $\Delta E = E_B - E_A$. All energies are electronic energies. $^{\#}$ Transition state. Guides to the number labels of the structures in the reactions can be found in Schemes 4 (*trans* pathway) and 5 (*cis* pathway).[a] Estimated by LST calculations. [b] Not available
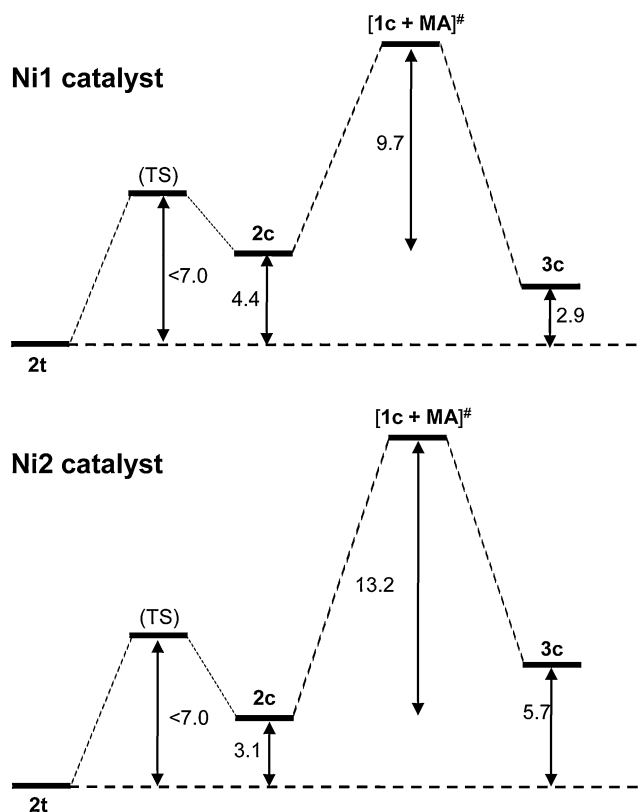
Fig. 1 Energy profiles for the isomerization steps with the Ni1 and Ni2 catalysts. Energies are given relative to those of the **2c** isomer in kcal mol^{-1}

the catalyst due to the formation of stable O-bound complexes (**2t**).

Different isomerization mechanisms were studied based on our previous results. In the first mechanism, we studied the isomerization from *trans*-O-bound complexes (**2t**) to the *cis*-2,1-π-vinyl (**3c**) in a single-step reaction (concerted mechanism). The LST calculations for this reaction (**2t** → **3c**) show high energy barriers for both catalysts (around 27.0 and 38.0 kcal mol^{-1} for Ni1 and Ni2, respectively).

An alternative isomerization mechanism was therefore proposed (Fig. 1). The first step in that mechanism is the *trans–cis* isomerization of the O-bound complex from the **2t** to the **2c** species. It was found that the **2t** species are 4.4 and 3.1 kcal mol^{-1} less stable than the **2c** ones for Ni1 and Ni2 catalysts, respectively. The corresponding energy barriers for this process according to LST calculations are estimated to be lower than 7.0 kcal mol^{-1} for both catalysts (Fig. 1).

After that, a new isomerization from these *cis* O-complexes (**2c**) to the *cis* 2,1-π-vinyl complexes (**3c**) can occur. A geometrical rearrangement where the MA becomes partially unbound from the metal center and changes from an O-interaction to a 2,1-π-vinyl complex takes place (**2c** → [**1c**–MA]$^{\#}$ → **3c**; see Fig. 1). The energy barriers for this step are estimated to be 9.7 and 13.2 kcal mol^{-1}, respectively, for Ni1 and Ni2. The energy differences between **3c** and **2c** isomers are −1.5 and +2.6 kcal mol^{-1} for Ni1 and Ni2 catalysts. Remarkably, the energy barrier from **2c** to **3c** is 3.5 kcal mol^{-1} lower for the Ni1 catalysts.

According to these results, **3c** isomers would be more feasible for the Ni1 catalyst than for the Ni2 one. The implications of these differences between the catalysts are considered in the "Discussion."

Methyl acrylate (MA) insertion after the isomerization process

Regarding the **3c** geometries, the π-bond of the MA monomer lies perpendicular to the Ni–Cα bond. As can be seen in Fig. 2, the Me-substituted catalyst presents a distorted version of the planar structure formed by the metal atom and the 2-iminopyridine-N-oxide ligand. The atoms in the Ni2_**3c** species remain practically in the same plane. This tendency is also observed in all of the structures formed during the copolymerization processes of both catalysts.

The MA insertion transition states **4c** are, respectively, 14.1 and 10.5 kcal mol^{-1} higher in energy than the corresponding π-complexes (**3c**) for Ni1 and Ni2 catalysts (see Table 1 and Scheme 5). The γ-agostic products **5c** are −2.0 and −5.6 kcal mol^{-1} more stable than the corresponding π-complexes (**3c**) for both catalysts. The related β-agostic products **6c** are −8.6 and −7.2 kcal mol^{-1} more stable than the γ-agostic ones (**5c**). These β-agostic products are the new active species that allow the polymerization process to continue, as will be presented in the following subsection.

Ethylene copolymerization and the formation of chelates

In order to continue with the polymerization process, a new ethylene monomer can be coordinated and inserted into the **6c** species (see Scheme 5). The coordination energies are −15.7 and −16.4 kcal mol^{-1} for Ni1 and Ni2 catalysts, respectively (see Table 1).

In both **8c** complexes, the π-bond of the MA is again perpendicular to the Ni–Cα bond. The insertion energy barriers are around 10 kcal mol^{-1} for the two catalysts. These energy barrier values indicate that both catalysts would be able to continue the polymerization process after a 2,1 MA insertion. For both catalysts, the direct insertion products **10c** are found to be six-membered chelates with an interaction between the inserted MA carbonyl O atom and the Ni atom (see Fig. 3). These products are −16.1 and −18.2 kcal mol^{-1} more stable than their corresponding π-complexes (**8c**) for Ni1 and Ni2, respectively.

Although these products are very stable, it was found that the complexation of a new ethylene monomer into these species leads to new π-complexes, **11c**, which are

Fig. 2 Geometries of the MA π-vinyl complexes for the Ni1 and Ni2 catalysts, showing a significant out-of-plane displacement of the Ni atom in the Ni1 case

Ni1_3c   Ni2_3c
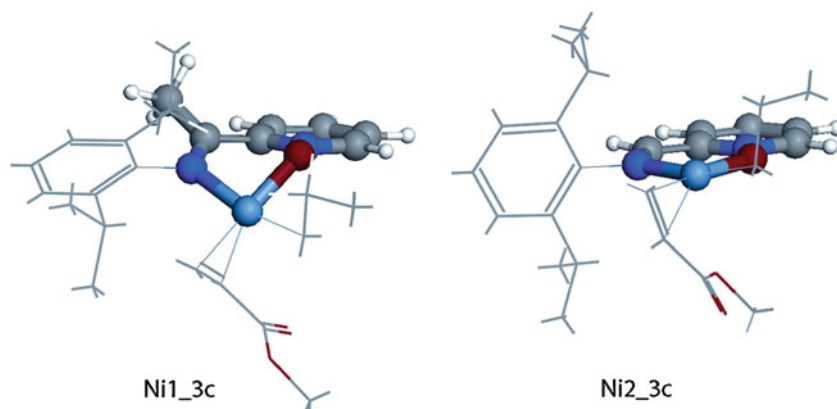
Scheme 5 Mechanism for MA polymerization along the *cis* pathway, as catalyzed by PymNox nickel catalysts
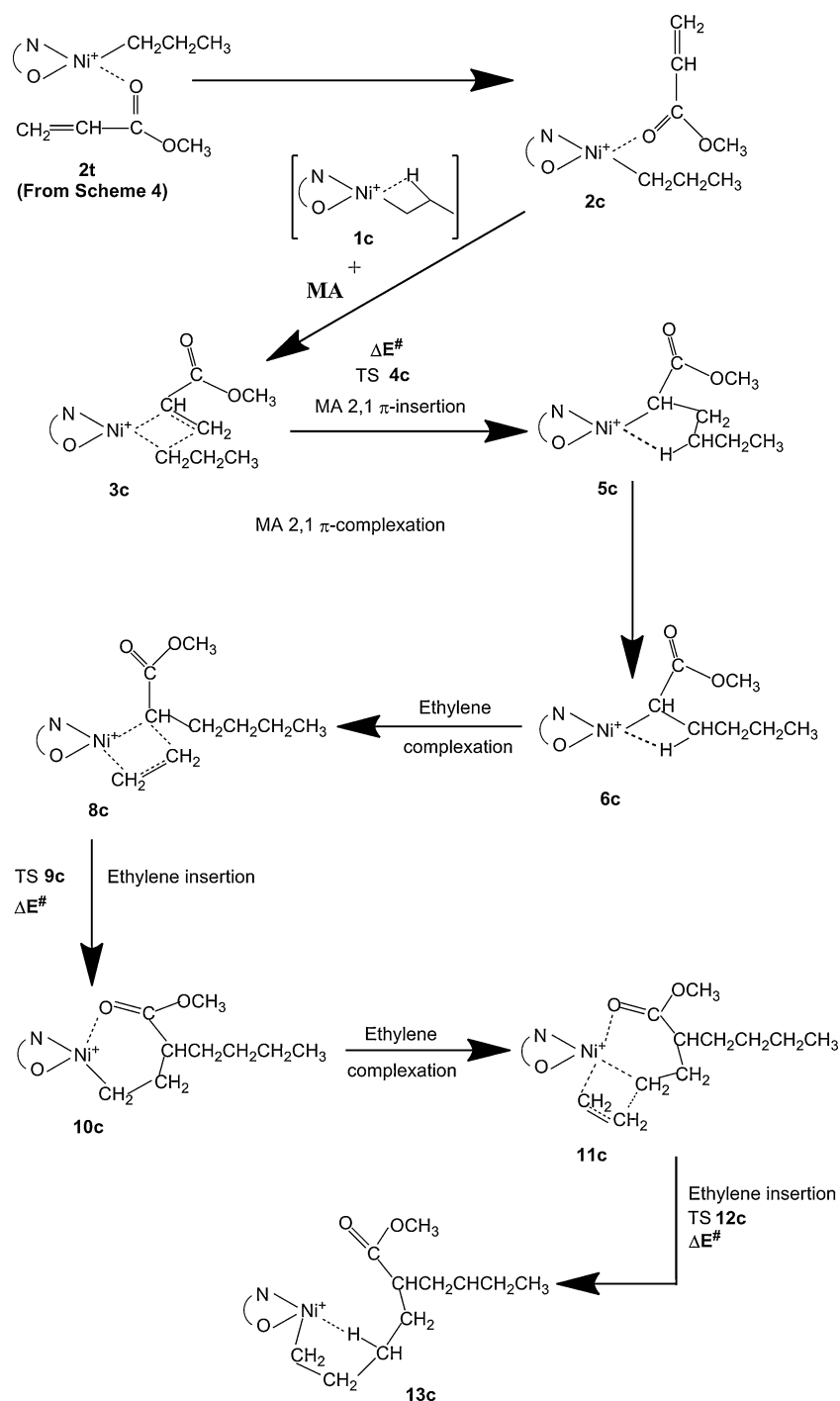
more stable than the **10c** species (see Table 1). In these complexes, trigonal bipyramidal geometries are found where the Ni atom has a coordination number of 5 (see the Ni1_**11c** and Ni2_**11c** geometries in the "Electronic supplementary material"). The five atoms involved in the bipyramid that surround the Ni atom are the N and O atoms of the PymNox ligand, the Cα atom of the polymer chain, and the two carbon atoms of the new ethylene molecule.

The energy barrier to the insertion of the ethylene monomer into these species has been estimated to be around 10 kcal mol^{-1}. The direct products of these new insertions (**13c**) present a γ-agostic interaction between the polymer chain and the metal atom, thus causing the square-planar geometry to be recovered. These products are −5.3 and 8.1 kcal mol^{-1} more stable than their corresponding π-complexes (**11c**) for Ni1 and Ni2 catalysts, respectively.
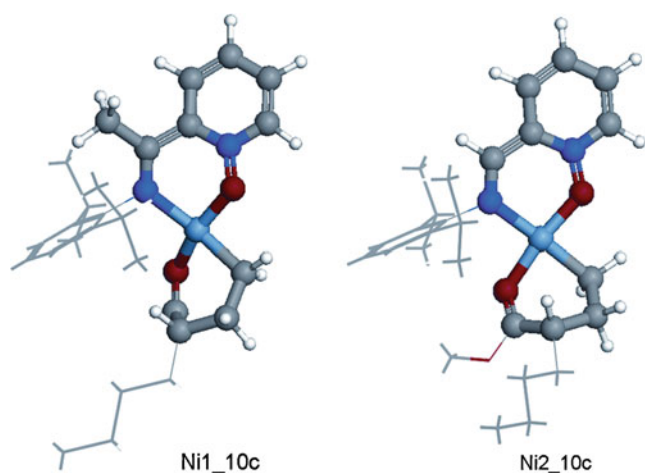
Fig. 3 Geometries of the six-membered ring chelates formed during the copolymerization process. **10c** shows pseudo-tetrahedral geometry around the metal atom for the Ni1 catalyst, whereas **10c** shows a planar coordination around the Ni atom for the Ni2 catalyst

Discussion

The species that can continue the polymerization process (resting states) during the ethylene–MA copolymerization reaction were found to be the **1t** geometries (Scheme 4). Scheme 3 shows the different modes for the coordination of both monomers into the catalysts. The O-bound coordination (**2t**) is more stable than the 2,1-π-vinyl MA coordination (**3t**) by 5.5 and 4.8 kcal mol^{-1} for Ni1 and Ni2 catalysts, respectively. This O-bound coordination (**2t**) is also more stable than ethylene-bound coordination by 1.3 and 2.0 kcal mol^{-1} for the Ni1 and Ni2 catalysts, respectively.

On the other hand, MA insertions from **3t** present high energy barriers when following the *trans* pathway (see Table 1). Similar results were obtained for the ethylene insertion monomer [37]. According to this, we postulate that insertion following the *trans* mechanism is not viable.

In this sense, isomerization processes from **2t** species were proposed in order to find a feasible path for the copolymerization process (Fig. 1). The equilibrium between **2t**, **2c** and **3c** isomers was explored. The rate-limiting barrier for this equilibrium is that for the **2c**↔**3c** step, which is 3.5 kcal mol^{-1} lower for the Ni1 catalyst than for the Ni2 one.

Furthermore, in the case of Ni1, the **3c** isomer is more stable than **2c** by −1.5 kcal mol^{-1}, while in the Ni2 catalyst, the difference in stability has the opposite sign: it is +2.6 kcal mol^{-1}. Thus, we suggest that the **3c** species are rarely reached in the case of the Ni2 catalyst, limiting the comonomer insertion.

MA insertion into **3c** for Ni1 catalyst can occur through a 14.1 kcal mol^{-1} energy barrier. At this point, the Curtin–

Hammett mechanism has been suggested, because escape from the **2t**↔**2c**↔**3c** equilibrium requires 10.6 kcal mol^{-1} less energy when it occurs through the *cis* pathway instead of the *trans* one (see Table 1 and energy profiles S1 and S2 in the "Electronic supplementary material").

Growing copolymer chains formed after consecutive insertions of MA/ethylene into Ni1 form six-membered chelate structures with an interaction between the Ni atom and the O atom of the polymer chain (Scheme 5 and Table 1). These structures are able to continue the polymerization process, as shown in "Ethylene copolymerization and the formation of chelates."

These differences between the catalysts can be related to the geometries of the optimized structures. It is apparent that MA monomer coordination in the Ni1 catalyst leads to distortion of the planar configuration around the Ni atom into a pseudo-tetrahedral environment. It is also observed that this distortion does not occur with the Ni2 catalyst, due to the lower steric hindrance of the unsubstituted ligand. The square-planar coordination persists around the Ni atom in the Ni2 catalyst (Cartesian coordinates for the geometries are given in the "Electronic supplementary material").

Conclusions

Computational studies of the cationic Ni-PymNox catalyst that is employed to copolymerize ethylene with MA comonomers as an alternative to the use of Grubbs' neutral catalysts are presented.

During the propagation chain step, the insertion takes always place along the *cis* path, mainly for two reasons: first, the high energy barriers found for ethylene and MA monomer insertions along the *trans* pathway (**3t** → [**4t**]$^{\#}$ → **5t**); second, the isomerization steps between intermediate complexes before insertions become feasible favor monomer insertions via the *cis* pathway. Thus, we propose that there is a Curtin–Hammett mechanism for the propagation steps during the copolymerization, rather than an Cossee–Arlman mechanism.

We have shown that the unsubstituted catalyst (Ni2) is less amenable to the incorporation of MA because the isomerization of the **2t** species to the **3c** isomer is much less favorable than it is for the substituted catalyst (Ni1). This finding is consistent with experimental observations.

The methyl-substituted catalyst is able to incorporate the MA monomer by coordinating with it through the π-system (**3c**). The MA insertion step may compete with olefin insertion. The steric effect exerted by the methyl group over the aryl ring, which in turn modifies steric congestion around the metal atom, appears to be the principal cause of

this different behavior. These results are able to explain the experimental observations of Campora et al. [20] that the methyl-substituted catalyst incorporates small amounts of MA as comonomer.

References

1. Santos JM, Ribeiro MR, Portela MF, Bordado JM (2001) Chem Eng Sci 56:4191–4196
2. Boffa LS, Novak BM (2000) Chem Rev 100:1479–1493
3. Imanishi Y, Naga N (2001) Prog Polym Sci 26:1147–1198
4. Yanjarappa MJ, Sivaram S (2002) Prog Polym Sci 27:1347–1398
5. Desurmont G, Tanaka M, Li Y, Yasuda H, Tokimitsu T, Tone S, Yanagase A (2000) J Polym Sci Polym Chem 38:4095–4109
6. Hagihara H, Tsuchihara K, Sugiyama J, Takeuchi K, Shiono T (2004) J Polym Sci A Polym Chem 42:5600–5607
7. Hakala K, Helaja T, Lofgren B (2000) J Polym Sci A Polym Chem 38:1966–1971
8. Kawahara N, Kojoh S, Matsuo S, Kaneko H, Matsugi T, Kashiwa N (2005) J Mol Catal A Chem 241:156–161
9. Wang YB, Hillmyer MA (2001) J Polym Sci A Polym Chem 39:2755–2766
10. Klabunde U, Mulhaupt R, Herskovitz T, Janowicz AH, Calabrese J, Ittel SD (1987) J Polym Sci A Polym Chem 25:1989–2003
11. Keim W, Appel R, Storeck A, Kruger C, Goddard R (1981) Angew Chem Int Ed 20:116–117
12. Klabunde U, Ittel SD (1987) J Mol Catal 41:123–134
13. Johnson LK, Killian CM, Brookhart M (1995) J Am Chem Soc 117:6414–6415
14. Stojcevic G, Prokopchuk EM, Baird MC (2005) J Organomet Chem 690:4349–4355
15. Williams BS, Leatherman MD, White OS, Brookhart M (2005) J Am Chem Soc 127:5132–5146
16. Connor EF, Younkin TR, Henderson JI, Hwang SJ, Grubbs RH, Roberts WP, Litzau JJ (2002) J Polym Sci A Polym Chem 40:2842–2854
17. Gibson VC, Tomov A (2001) Chem Commun 1964–1965
18. Younkin TR, Conner EF, Henderson JI, Friedrich SK, Grubbs RH, Bansleben DA (2000) Science 287:460–462
19. Rojas RS, Galland GB, Wu G, Bazan GC (2007) Organometallics 26:5339–5345
20. Brasse M, Campora J, Palma P, Álvarez E, Cruz V, Ramos J, Reyes ML (2008) Organometallics 27:4711–4723
21. Cruz V, Ramos J, Munoz-Escalona A, Lafuente P, Pena B, Martinez-Salazar J (2004) Polymer 45:2061–2072
22. Cruz VL, Martinez J, Martinez-Salazar J, Ramos J, Reyes ML, Toro-Labbe A, Gutierrez-Giliva S (2007) Polymer 48:7672–7678
23. Exposito MT, Martinez S, Ramos J, Cruz V, Lopez A, Munoz-Escalona A, Haider N, Martinez-Salazar J (2004) Polymer 45:9029–9038
24. Martinez S, Exposito MT, Ramos J, Cruz V, Martinez MC, Lopez M, Munoz-Escalona A, Martinez-Salazar J (2005) J Polym Sci A Polym Chem 43:711–725
25. Ramos J, Munoz-Escalona A, Martinez S, Martinez-Salazar J, Cruz V (2005) J Chem Phys 122(74901):1–4
26. Cruz VL, Ramos J, Martinez S, Munoz-Escalona A, Martinez-Salazar J (2005) Organometallics 24:5095–5102
27. Ramos J, Cruz V, Munoz-Escalona A, Martinez-Salazar J (2001) Polymer 42:8019–8023
28. Musaev DG, Morokuma K (1999) Top Catal 7:107–123
29. Michalak A, Ziegler T (2006) Kinet Catal 47:310–325
30. Ramos J, Cruz V, Munoz-Escalona A, Martinez S, Martinez-Salazar J (2003) Polymer 44:2169–2176
31. Ramos J, Munoz-Escalona A, Cruz V, Martinez-Salazar J (2003) Polymer 44:2177–2186
32. Michalak A, Ziegler T (2002) J Am Chem Soc 124:7519–7528
33. Michalak A, Ziegler T (2001) Organometallics 20:1521–1532
34. Deubel DV, Ziegler T (2002) Organometallics 21:1603–1611
35. Deubel DV, Ziegler T (2002) Organometallics 21:4432–4441
36. Michalak A, Ziegler T (2003) Organometallics 22:2660–2669
37. Ramos J, Cruz VL, Martinez-Salazar J, Brasse M, Palma P, Campora J (2010) J Polym Sci A Polym Chem 48:1160–1165
38. Te Velde G, Bickelhaupt FM, Baerends EJ, Fonseca Guerra C, van Gisbergen SJA, Snijders JG, Ziegler T (2001) J Comput Chem 22:931–967
39. Vosko SH, Wilk L, Nusair M (1980) Can J Phys 58:1200–1211
40. Perdew JP, Yue W (1986) Phys Rev B 33:8800–8802
41. Becke AD (1998) Phys Rev A 38:3098–3100

ORIGINAL PAPER

Insight into the reaction mechanism of *cis,cis*-muconate lactonizing enzymes: a DFT QM/MM study

Tuanjai Somboon · Matthew Paul Gleeson ·
Supa Hannongbua

Abstract MLEs derived from *mycobacterium smegmatis* and *seudomonas fluorescens* share ~76% identity and have a very similar arrangement of catalytic residues in their active site configuration. However, while they catalyze the conversion of *cis,cis*-muconate to the same achiral product, muconolactone, studies in deuterated solvent surprisingly show that the cyclo-isomerization proceeds with the formation of a chiral product. In this paper we discuss the application of DFT QM/MM calculations on both MLEs, to our knowledge the first reported in the literature on this protein. We investigate the proposal that the base involved in the catalytic reaction is the lysine residue found at the end of the 2nd strand given: (a) that the lysine residue at the end of the 6th strand is in an apparently equally effective position to catalyze reaction and (b) that the structural related epimerase in-fact achieve their stereo-specific outcomes by relying on either the base from the 2nd or 6th strand.

T. Somboon · M. P. Gleeson (✉) · S. Hannongbua
Department of Chemistry, Faculty of Science,
Kasetsart University,
50 Phaholyothin Road, Chatuchak,
Bangkok 10900, Thailand
e-mail: paul.gleeson@ku.ac.th

S. Hannongbua (✉)
Center of Nanotechnology KU, Kasetsart University,
50 Phaholyothin Road, Chatuchak,
Bangkok 10900, Thailand
e-mail: fscisph@ku.ac.th

Introduction

The enolase superfamily has received a considerable amount of attention from a biochemical perspective recently as it has helped to illustrate the complexity and redundancy in enzyme evolution [1–3]. Muconate lactonizing enzymes (MLEs) are interesting members of this superfamily which catalyze the conversion of *cis,cis*-muconates to muconolactones. Sakai et al. [4–6] have recently reported data on two different MLEs derived from *mycobacterium smegmatis* and *seudomonas fluorescens*, that share ~76% identity. These proteins catalyze the same chemical reaction (Scheme 1), but involve a stereochemically distinct reaction mechanism even though the product is achiral. Studies in deuterated solvent have established that cyclo-isomerization catalyzed by *mycobacterium smegmatis* (*anti*-MLE) proceeds along an *anti*-stereochemical course, whereas *seudomonas fluorescens* (*syn*-MLE) catalyzes the *syn*-stereochemical course.

MLEs are Mg^{2+} containing metallo-proteins consisted of ~370 amino acids arranged into a TIM barrel-like α/β protein fold. The Mg^{2+} is found deep within a non-solvent exposed cavity, coordinated by 2 aspartate and 1 glutamic acid residues, a single water molecule and the substrate, in a distorted octahedral form. The carboxylate of *cis,cis*-muconate binds across the Mg^{2+} ion in conformations that display mirror-like symmetry in the two different proteins (Fig. 1). The 2nd carboxylate group of *cis,cis*-muconate interacts with residues toward the rear of the pocket, glutamine and threonine in *anti*-MLE and histidine and threonine residues in *syn*-MLE. Two Lys residues, located on the 2nd and 6th

Scheme 1 Reaction mechanism proposed by Sakai et al. for *anti*- and *syn*-MLEs. Two proximal Lys residues exist in both active sites, which could potentially lead to the formation of either the *syn*- or *anti*-products. In fact, while both proteins catalyze the formation of the same natural precursor, they lead to the formation of different chiral products based on experiments performed in deuterated solvent. It is also unclear whether the enolate intermediate proposed by Sakai et al. is a stable intermediate during the reaction

strands, are sufficiently closed to the substrate alpha carbon to act as the general base in the reaction.

Structural information has proved crucial to understand the sequence of events that lead to the chiral products in deuterated solvent. Sakai et al. have determined X-ray crystal structures of the product state (muconate lactone), for both *syn*- and *anti*-proteins, and have used these to identify the most probable base in the catalytic reaction. They propose that the identity of the base in both proteins is the Lys residue at the end of the 2nd strand that explains the stereochemical aspects of the reaction. This result is interesting given members of the structurally related epimerase sub-class achieve their stereo-specific outcomes by relying on either the base from the 2nd or 6th strand. Furthermore, analysis of the interaction distances between the nitrogen atoms of the two possible bases and the product α-carbon in both MLE PDBs (Table 1) simplistically suggests that Lys on the 6th strand is more likely to be the base in *syn*-MLE (3.03Å vs 3.57Å). This however would not explain the stereo-chemical differences in the reaction. Additionally, it is not known if a stable enolate anion does in fact exist as a meta-stable intermediate in the reaction.

Computational chemistry can play an important role in understanding protein function since it enables us to simulate events at an atomic level. In this study we apply hybrid quantum mechanics/molecular mechanics (QM/MM) calculations to probe aspects of MLE function and gain further insight into the catalytic events in both *anti*- and *syn*-MLEs that lead to the observed stereochemical differences. In the QM/MM technique the active site residues that undergo chemical change, or directly influence the sequence of events in the catalytic reaction, are treated using more accurate QM methods, while the remainder of the protein is treated using less rigorous, but more computationally efficient MM methods. The QM/MM calculations performed here rely on the ONIOM methodology [7, 8] using the electrical embedding scheme. Here the total QM/MM energy of the system is computed in a subtractive fashion as given in the following equation; QM energy of the active site region, or "model", plus the MM energy of the "real" or whole protein system, minus the MM energy of the model region. For more extensive reviews of the QM/MM technique see the following references [9–12].

$$E_{QM/MM} = E_{QM-EE(model)} + E_{MM(real)} - E_{MM(model)} \qquad (1)$$

QM/MM methods have been applied to study the enolase family member phosphoenolpyruvate. This protein contains two Mg^{2+} ions in the active site and the mechanism involves the two separate CH proton abstraction steps. Liu et al. [13] reported QM/MM free energy perturbation barrier heights of ~13.1 kcal mol^{-1} for the initial proton abstraction step, resulting in a stable enolate intermediate of ~5 kcal mol^{-1}. Decomposition of the intermediate by abstraction of a further proton from the β-carbon was found to require ~9 kcal mol^{-1}. The stability of enolate intermediates has also been investigated by Van Der Kamp et al. [14] in the unrelated protein citrate synthase. They find that proton abstraction from the α-carbon of oxaloacetate requires ~10.2 kcal mol^{-1} and results in a stable intermediate ~8 kcal mol^{-1} higher in energy than

Fig. 1 A comparison of the *anti*- (blue) and *syn*-MLE (green) proteins (left) and active site regions
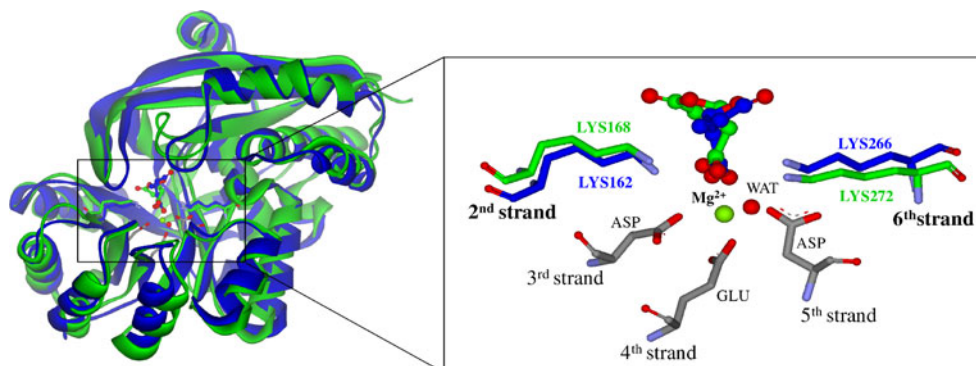
Table 1 Predicted and experimental distances observed for the conversion of *cis,cis*-muconate to muconolactone during the MLE catalyzed reactions. Distances are reported in Å

Distances	Structures					
Anti-MLE	Reactant	*Syn*-TS	*Syn*-Product	*Anti*-TS	*Anti*-Product	X-ray (product)
Mg−O(ASP191)	2.35	2.47	2.30	2.41	2.22	2.08
Mg−O(GLU217)	2.02	2.02	2.00	2.01	2.02	1.96
Mg−O(ASP242)	2.00	2.01	1.95	2.02	2.05	2.19
Mg−O(Sub)	2.31	2.29	2.23	2.17	2.23	2.16
Mg−O(Sub)	2.21	2.12	2.49	2.22	2.36	2.29
Mg−O(WAT)	2.06	2.06	2.10	1.99	1.99	2.20
C(Sub)−N(LYS162)	3.11	3.39	3.45	2.99	3.04	3.03
C(Sub)−N(LYS266)	3.66	3.03	3.87	3.56	3.66	3.57
C(Sub)−HN(LYS162)	2.25	2.47	2.67	1.90	1.10	-
C(Sub)−HN(LYS266)	2.92	2.10	1.09	2.81	2.90	-
O(Sub)−H(GLN294)	1.86	1.90	2.01	1.84	1.88	-
Syn-MLE	Reactant	*Syn*-TS	*Syn*-Product	*Anti*-TS	*Anti*-Product	X-ray (Product)
Mg−O(ASP197)	2.59	2.62	2.27	2.72	2.65	2.34
Mg−O(GLU223)	1.99	2.02	2.01	2.02	2.00	2.33
Mg−O(ASP248)	2.04	2.03	2.10	2.02	1.95	2.24
Mg−O(Sub)	2.15	2.07	2.16	2.16	2.03	2.42
Mg−O(Sub)	2.22	2.22	2.49	2.11	3.37	2.96
Mg−O(WAT)	1.99	2.00	1.99	1.98	1.97	1.79
C(Sub)−N(LYS168)	3.66	3.07	3.60	4.24	4.26	4.22
C(Sub)−N(LYS272)	3.55	4.02	3.81	2.72	3.31	3.45
C(Sub)−HN(LYS168)	2.80	2.00	1.09	3.33	3.31	-
C(Sub)−HN(LYS272)	3.13	3.77	3.42	1.93	1.10	-
O(Sub)−H(HIE21)	1.78	1.86	1.88	1.89	1.82	-

the reactants. While the mechanisms differ considerably from MLE, the results discussed above serve as useful benchmarks for studies on MLEs.

In this paper we discuss the application of QM/MM calculations to MLEs, to our knowledge the first reported in the literature on this system. The goal of this study is to help conclusively prove the identity of the base in the *anti*- and *syn*-MLE proteins given the contrasting way related enolase family member achieve stereo-specific catalysis and the apparent ambiguities from an analysis of the X-ray interactions distances between possible bases. We are also interested in assessing whether an explicit enolate anion exists as a reaction intermediate as found in other enolases, or as a transition state.

Computational details

The crystal structures of *anti*-MLE (pdb accession code: 3DG6) and *syn*-MLE (pdb accession code: 3DGB) were downloaded from RCSB protein data bank and prepared as follows. Cofactors, ions, and water molecules beyond 15 Å

of the active site were deleted. Missing side chain data from both PDB structures as well as hydrogen atoms were added using Discovery Studio 2.5 according to the CHARMM 22 forcefield. The protonation states of ionizable residues were determined by visual analysis. Ligand charges were determined using the AM1BCC method and parameters according to the Accelrys CHARMm forcefield. Both proteins were solvated in a box of TIP3P water with a minimum distance of 7 Å between the protein and box edge (i.e, 14 Å between proteins in a periodic box). Counterions were added to neutralize the system. Default non-bonded cut-offs of 12 Å were used in all MM simulations.

Due to the difficulty in accurately simulating metalloproteins using MM methods, the Mg^{2+} ion, its three chelating carboxylate groups and one water molecule were harmonically restrained to their X-ray positions during all of the MM preparation steps. MM optimization was achieved in three distinct steps. All optimizations were performed in Discovery Studio 2.5 [15] using the smart optimizer conditions and an RMS gradient below 0.1 kcal mol^{-1}. These were; (1) optimization of hydrogen atoms only followed by (2) optimization of all amino acid side chain

atoms and solvent molecules and finally (3) optimization of all atomic coordinates. This setup is equivalent to the default protein preparation procedure to prepare protein X-ray structures for docking, molecular dynamics or QM/MM in the modeling package Maestro [16].

The MM optimized coordinates were then used in a subsequent short molecular dynamics (MD) step to help minimize any high energy contacts that are often present in X-ray protein structures. Atoms beyond 10 Å of the active site were harmonically restrained. MD was performed in two stages; (a) heating from T=0 to 300K over 200 ps. (b) equilibration for 800 ps. Simulations were performed using the default CHARMM settings in Discovery Studio 2.5. These stages include a time step of 0.001 ps., NVT conditions, 12 Å non-bonded cut-offs and particle mesh Ewald [17]. The flexible atoms from the final MD step were subsequently re-optimized and used as input for QM/MM calculations.

All QM/MM calculations were performed using the ONIOM methodology developed by Morokuma and co-workers as implemented in Gaussian 03 [18]. A QM region has been selected so that key polar residues that directly interact with the substrate over the course of the reaction are included explicitly. For anti-MLE 70 atoms are treated QM consisting of the side chains of; SER23, THR54, LYS162, ASP191, GLU217, ASP242, LYS266, GLN294, the Mg^{2+} ion and the substrate. For syn-MLE 76 atoms are treated QM consisting of the side chains of; HIE21, THR140, LYS168, ASP197, GLU223, ASP248, LYS272, the Mg^{2+} ion and the substrate. The side chains of the following active site residues were treated flexibly with the rest of the system being fixed (anti-MLE: PHE21, PHE53, LYS160, ASN193, ILE295; syn-MLE: ILE53, THR58, LYS166, ASN199, THR300, LEU302, GLU326, PHE328). All water molecules were removed for computational efficiency except the water molecule that chelates Mg^{2+}.

The M05 functional developed by Zhao et al. [19, 20] has been used in these calculations as it has been shown to be more effective than the popular B3LYP method for describing aspects of non-bonded interactions. The 6-31G(d) basis set was employed for geometry optimizations. Single point energies on stationary points being characterized using the 6-311+G(d,p) basis set. The MM region was treated using universal Force field (UFF) in conjunction with CHARMM partial charges. The reaction coordinates of both possible bases, in both proteins, have been estimated by the stepwise variation of the C-O and C-H bonds between their reactant and product configurations. Due to the large memory requirements of Gaussian 03 with MM regions of this size (>5000 atoms) transitions states were characterized as saddle points by doing a frequency calculations of the optimized QM region

coordinates only. ONIOM optimization has been performed using default settings; fixed link atom positions and involves the electrical embedding of MM charges into the QM calculation.

Results and discussion

In our attempt to confirm the identity of the catalytic base in the anti- and syn-MLE proteins, we have determined the minimum energy structures for the reactants and both the anti- and syn-products in both proteins. As the products may be dictated by kinetic factors, we have also determined the potential energy surface between reactant and products to estimate the reaction barriers and also to assess the possibility of an explicit enolate anion existing during the course of the reaction.

X-ray versus QM/MM structures

Firstly we analyze the differences between the two different product structures for each protein, which differ only in terms of the position of a single proton, and compare these to the corresponding heavy atom coordinates reported in the respective X-ray structure PDB files. The expectation is that the most thermodynamically favorable QM/MM coordinates will be more similar to the original X-ray structure and that the corresponding base will give rise to the expected stereochemical outcome.

Analysis of the RMSDs of the QM/MM optimized active site regions of the two possible products could potentially be used to identify which base is involved in the catalytic reaction. One would expect that the optimized product that displays the lowest RMSD to the experimental X-ray structure would identify which base is involved in the reaction. Listed in Table 1 are the key geometrical parameters associated with the stationary points. While the optimized QM/MM structures of both products for both anti- and syn-MLE proteins, display C_α RMSDs of <0.05 Å to the corresponding X-ray structure, the flexible QM and MM heavy atoms in the QM/MM system understandably display larger differences in structure. Note, supplementary information Fig. S1 displays the overlay of (a) the original X-ray coordinates, (b) the MD output structure and (c) the QM/MM optimized geometries, highlighting the rather small differences between them. For syn-MLE we find that the product structure involving the base which leads to the syn-product (Lys-168 located on the 2nd strand) has a lower RMSD than that formed with Lys-272, which results in the anti-product (RMSDs of 0.36 vs 0.56 Å). Furthermore, we also find that the anti-product formed within the anti-MLE active site has a lower overall RMSD that that of the syn-

product (0.58 vs 0.63 Å) suggesting that the 2nd strand lysine, Lys-162, is the base in this reaction and not Lys-266 located on the 6th strand.

Looking in more detail at the product structures (Table 1) we observe that the key interacts made in the *anti-anti* product are clearly in better agreement with the X-ray coordinates than the *anti-syn* product. However, for *syn*-MLE both products show distances considerably more variable from the X-ray structure. In fact, the configuration of the Mg^{2+} ion and its ligands in the *syn*-MLE QM/MM models is more similar to that observed for the *anti*-MLE products. It is possible that this artifact may have arisen from the fitting of the atomic solution of *syn*-MLE to its electron density. In fact, the Mg-O distances observed in the *syn*-MLE X-ray structure appear considerably longer than the experimentally expected values of ~2.1 Å [21]. Contrasting the two experimental *anti*- and *syn*-MLE X-ray structures (Table 1 and Fig. 1) we can see that Mg^{2+} coordination in the *syn*-MLE structure is much weaker than in the *anti*-MLE. Interestingly, the QM/MM optimized *syn*-product of *syn*-MLE display interactions that are more comparable to the experimental *anti*-product.

Assessment of the relative QM/MM energies shows that the *anti*-MLE *anti*-product (*anti-anti* product) is lower than the corresponding *anti*-MLE *syn*-product (*anti-syn* product) by 26.1 kcal mol^{-1}. Thus, from a purely thermodynamic perspective we would expect Lys-162 to act as the catalytic base since it leads to the lowest energy product and will lead to the known stereochemical outcome. In addition, the energetic differences between the *syn-syn* product and *syn-anti* product shows that the former is favored by 19.8 kcal mol^{-1}. Thus, from a purely thermodynamic perspective, we would expect Lys-168 to act as the catalytic base since it leads to the lowest energy product and the known stereochemical product.

To try and decipher the contribution electrostatics and van der Waals interactions, as well as protein pre-organization, we subsequently performed gas phase single point QM calculations on each of the QM/MM optimized product geometries. Analysis of the contributing terms to the QM/MM energy reveals that the Van der Waals term is essentially constant for structures obtained in the two different protein models. The gas phase QM single point energies show that both the *syn-syn* and *anti-anti* products are still preferred in the gas phase suggesting active site pre-organization is important (ΔE between *syn-syn* and *syn-anti* products are −26 vs −33 kcal mol^{-1} at QM/MM and QM levels respectively; ΔE between *anti-anti* and *anti-syn* products are −20 vs −6 kcal mol^{-1} at QM/MM and QM levels respectively). Interestingly, the inclusion of the protein electrostatic term has a much greater effect in stabilizing the *anti-anti* product over the *anti-syn* product since the difference in QM energy is just 6 kcal mol^{-1} in the

gas phase calculation but 20 kcal mol^{-1} in the protein. In contrast, the difference in energy between the *syn-syn* and *syn-anti* product is somewhat larger in the gas phase than in the protein calculations. This might suggests that the two proteins achieve their product selectivities in subtly different ways, *anti*-MLE relying more on organizing its active site to favor the *anti*-product conformation and *syn*-MLE by preferentially stabilizing the *syn*-product as a result of its particular electrostatic characteristics.

QM/MM reaction energetics

The QM/MM calculations performed here on the two possible products formed by both *anti*- and *syn*-MLEs appear to confirm the identity of the catalytic bases in the reaction from both a structural and energetic perspective. However, it is also desirable to assess the reaction barriers associated with the two possible bases in each protein to rule out any differences that might affect the reaction products from a kinetic perspective. We have therefore mapped out the potential energy surface between reactants and both possible products in *anti*- and *syn*-MLE proteins. From this we have estimated the reaction barriers for each process and these results are summarized in Fig. 2. The full potential energy surfaces for the *syn-syn* and *anti-anti* processes are summarized in Fig. 3.

The overall QM/MM results summarized in Fig. 2 show that the barriers to reaction associated with the *anti-anti* product is 11.6 kcal mol^{-1} lower than the corresponding *anti-syn* product. Similarly, we observe a 13.0 kcal mol^{-1} difference between the *syn-syn* and *syn-anti* products of *syn*-MLE. These results clearly suggest that the basic residues involved in the reaction are located at the end of the 2nd strand of both *anti*-MLE and *syn*-MLE, corresponding to Lys-162 and Lys-168 respectively. From
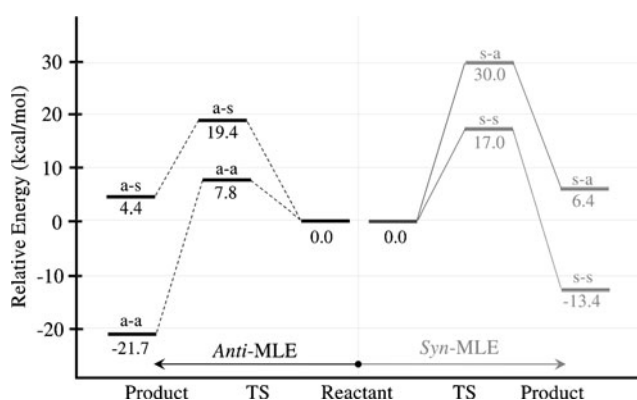

Fig. 2 Relative M05/6-31G(d):UFF optimized energies for the *ant*- and *syn*-products in both *anti*- and *syn*-MLEs. a-a refers to *anti*-product and a-s refers to *syn*-product of *anti*-MLE. For *syn*-MLE, s-s refers to *syn*-product and s-a refers to *anti*-product
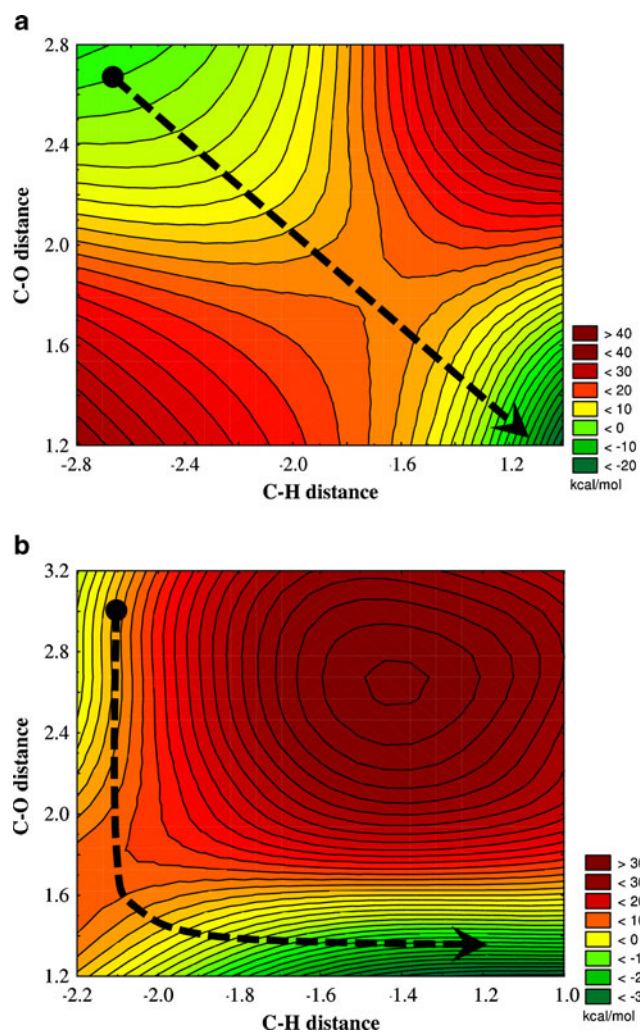
Fig. 3 Potential energy surfaces for the *anti-anti* (left) and *syn-syn* catalyzed reactions. In the former, P162 acts as the catalytic acid and in the latter, P168. C-O and C-H distances correspond to d_1 and d_2 in scheme 1, respectively. Energies are at the M05/6-31G(d):UFF level of theory. The energy performed at M05/6-31G(d) is correlated well with M05/6-311+G(d,p)

an analysis of Fig. 3 it is also apparent that the reaction proceeds along a concerted pathway with both C-O bond breaking and proton transfer to Lys-162. In contrast the process in *syn*-MLE appears to proceed along a more stepwise route, with C-O bond breaking occurring before proton transfer to Lys-168.

The theoretical QM/MM DFT models employed have helped to confirm the identity of the base in the reaction using. We were also therefore interested in understanding whether an explicit enolate anion exists in the active site of either protein. To determine whether the enolate was a stable stationary point within the protein we took the relevant structure obtained from the QM/MM potential energy scans (i.e, the structure having a C-O bond formed but a long C-H bond). This constrained structure was

subsequently fully optimized QM/MM using the conditions as used for the other stationary points obtained in this study. In the case of both *syn*- and *anti*-MLE this enolate-like structure decomposed to the muconolactone by accepting a proton from the active site lysine residue which leads to the forming of the C-O bond. While this result would appear to suggest that an explicit enolate anion does not exist in these proteins further work is needed to prove whether this high energy structure is truly a stable stationary point or not. These calculations would necessitate additional polarization and diffuse functions as well as full frequency analyses given that the structure is likely to occupy a shallow energy minimum.

Conclusions

In this paper we discuss the application of DFT QM/MM calculations on *cis,cis*-muconate lactonizing enzymes, to our knowledge the first reported in the literature. QM/MM methods have been used to determine the reaction energetics associated with the conversion of *cis,cis*-muconate to muconolactone in both *anti*- and *syn*-MLEs. We have investigated the proposal that the base involved in the catalytic reaction is the Lys residue found at the end of the 2nd strand, rather than a Lys residue 6th which is almost equally well positioned.

Our QM/MM results show that the expected muconolactone *anti*-product derived from *anti*-MLE is (a) closer to the X-ray structure in terms of RMSD, (b) lower in energy and (c) has a lower barrier to reaction than the corresponding *syn*-product. Our results also show that the *syn*-product derived from *syn*-MLE also displays a lower RMSD to the original X-ray coordinates, is lower in energy, and has a lower barrier to reaction than the corresponding *anti*-product. Thus, for *anti*-MLE we find Lys-162 to results in the lowest energy reaction and Lys-168 for *syn*-MLE.

The theoretical calculation performed here have helped to confirm the identity of the basic residues involved in the MLE reaction originally proposed by Sakai et al. through indirect means from an analysis of the products produced in experiments performed in deuterated solvents. Our results show that although the basic Lys residues located on the 6th strand in both MLEs are almost equally well positioned to accept a proton, however, it is the base located on the 2nd strand that is thermodynamically and kinetically more favorable, as well as giving the optimized QM/MM product closest to the original X-ray structure.

References

1. Gerlt JA, Babbitt PC, Rayment I (2005) Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. Arch Biochem Biophys 433:59–70. doi:10.1016/j.abb.2004.07.034

2. Thompson TB, Garrett JB, Taylor EA, Meganathan R, Gerlt JA, Rayment I (2000) Evolution of enzymatic activity in the enolase superfamily: structure of o-succinylbenzoate synthase from escherichia coli in complex with Mg2+ and o-succinylbenzoate†,‡. Biochemistry 39:10662–10676. doi:10.1021/bi000855o

3. Vick JE, Schmidt DMZ, Gerlt JA (2005) Evolutionary potential of (β/α)8-Barrels: In vitro enhancement of a "new" reaction in the enolase superfamily†. Biochemistry 44:11722–11729. doi:10.1021/bi050963g

4. Sakai A, Fedorov AA, Fedorov EV, Schnoes AM, Glasner ME, Brown S, Rutter ME, Bain K, Chang S, Gheyi T, Sauder JM, Burley SK, Babbitt PC, Almo SC, Gerlt JA (2009) Evolution of enzymatic activities in the enolase superfamily: stereochemically distinct mechanisms in two families of cis, cis-muconate lactonizing enzymes†‡. Biochemistry 48:1445–1453. doi:10.1021/bi802277h

5. Sakai A, Fedorov AA, Fedorov EV, Schnoes AM, Glasner ME, Brown S, Rutter ME, Bain K, Chang S, Gheyi T, Sauder JM, Burley SK, Babbitt PC, Almo SC, Gerlt JA (2009) Evolution of enzymatic activities in the enolase superfamily: stereochemically distinct mechanisms in two families of cis, cis-muconate lactonizing enzymes. Biochemistry 48:2569–2570. doi:10.1021/bi900265w

6. Sakai A, Xiang DF, Xu C, Song L, Yew WS, Raushel FM, Gerlt JA (2006) Evolution of enzymatic activities in the enolase superfamily: N-succinylamino acid racemase and a new pathway for the irreversible conversion of d- to l-amino acids†. Biochemistry 45:4455–4462. doi:10.1021/bi060230b

7. Dapprich S, Komáromi I, Byun KS, Morokuma K, Frisch MJ (1999) A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. J Mol Struct THEOCHEM 461–462:1–21. doi:10.1016/s0166-1280(98)00475-8

8. Vreven T, Byun KS, Komáromi I, Dapprich S, Montgomery JA, Morokuma K, Frisch MJ (2006) Combining quantum mechanics methods with molecular mechanics methods in ONIOM. J Chem Theor Comput 2:815–826. doi:10.1021/ct050289g

9. Bruice TC (2006) Computational approaches: reaction trajectories, structures, and atomic motions. Enzyme reactions and proficiency. Chem Rev (Washington, DC US) 106:3119–3139. doi:10.1021/cr050283j

10. Lin H, Truhlar DG (2007) QM/MM: What have we learned, where are we, and where do we go from here? Chem Inform. doi:10.1002/chin.200722224

11. Senn HM, Thiel W (2007) Atomistic approaches in modern biology. In: Topics in current chemistry, vol 268. Springer, Berlin

12. Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. Angew Chem Int Edn 48:1198–1229. doi:10.1002/anie.200802019

13. Liu H, Zhang Y, Yang W (2000) How is the active site of enolase organized to catalyze two different reaction steps? J Am Chem Soc 122:6560–6570. doi:10.1021/ja9936619

14. Kamp MWvd, Perruccio F, Mulholland AJ (2008) High-level QM/MM modelling predicts an arginine as the acid in the condensation reaction catalysed by citrate synthase. Chem Commun (Cambridge, UK) 16:1874–1876

15. Drug Discovery Studio 2.5 (2010) Accelrys Inc, San Diego, CA, USA

16. Maestro, version 9.1 (2010) Schrödinger, LLC, New Yok, NY

17. Darden TYD, Pedersen L (1993) Particle mesh Ewald: An N•log (N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092

18. 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, 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 J (2004) Gaussian 03, Revision B.05. Gaussian Inc, Wallingford, CT

19. Zhao Y, Truhlar DG (2006) Density functionals for noncovalent interaction energies of biological importance. J Chem Theor Comput 3:289–300. doi:10.1021/ct6002719

20. Zhao Y, Truhlar DG (2008) Density functionals with broad applicability in chemistry. Acc Chem Res 41:157–167. doi:10.1021/ar700111a

21. Plyasova LM, Vasilieva NA, Cherepanova SV, Shmakov AN, Chuvilin AL (1998) Structure investigation of defect MgO - high temperature process catalyst. Nucl Instrum Meth Phys Res A 405:473–475. doi:10.1016/s0168-9002(97)00161-7

Molecular dynamics simulations of nafion and sulfonated poly ether sulfone membranes II. Dynamic properties of water and hydronium

**Takahiro Ohkubo · Koh Kidena · Naohiko Takimoto ·
Akihiro Ohira**

Abstract We measured the self-diffusion coefficients of water in a Nafion membrane and two sulfonated polyethersulfone (SPES) membranes with varying ion-exchange capacities (IEC) in terms of relative humidity using the pulse field gradient NMR (PFG-NMR) technique. The self-diffusion coefficients were plotted against the number of water molecules per sulfonic acid group, λ, and compare these values with the results of molecular dynamics (MD) simulations. Classical MD simulations for all membranes were carried out using a consistent force field at $\lambda=3$, 6, 9, 12, and 15. The dynamic properties of water (H_2O) and hydronium (H_3O^+) on a molecular level were estimated as self-diffusion coefficients and residence times around a sulfonate group (SO_3^-). The diffusion coefficients of H_2O and H_3O^+ followed the order, Nafion>SPES with IEC=1.4 >SPES with IEC=1.0>SPES with IEC=0.75, which agreed with the experimental data. The residence time distribution of H_2O around SO_3^- in Nafion was in the range of 1–6 ps, whereas H_2O in the SPES exhibited a residence time of greater than 20 ps.

Keywords Molecular dynamics · Nafion membrane ·
Sulfonated polyether sulfone membrane · Proton exchange
membrane fuel cell · Aqueous phase structure

T. Ohkubo (✉)
Applied Chemistry and Biotechnology,
Graduate School of Engineering, Chiba University,
1-33 Yayoi-cho Inage-ku,
Chiba 263–8522, Japan
e-mail: ohkubo.takahiro@faculty.chiba-u.jp

K. Kidena · N. Takimoto · A. Ohira
Fuel Cell Cutting-Edge Center (FC-Cubic),
Technology Research Association,
2-3-26 Aomi, Koto-ku,
Tokyo 135–0064, Japan

Introduction

Proton-exchange-membrane fuel cells (PEMFC) convert chemical energy to electrical energy with high efficiency and minimal pollution. Because the PEM for conducting protons between electrodes is a critical component for the performance of PEMFC, it is essential to understand the dynamic properties of protons and the fundamental process of proton transport in the membrane.

Several membranes have been considered for practical use in PEMFCs. Proton conductivity and water diffusion in hydrated Nafion have been examined by different kinds of experimental approach. Zawodzinski et al. [1] determined proton diffusion coefficients from conductivity measurements and proton diffusion coefficients from pulse field gradient NMR (PFG-NMR) in Nafion at various levels of hydration (λ ranging from 2 to 22). These authors suggested that structural diffusion without displacement of protons is enhanced at higher λ because of better connectivity of the water channel. Pivovar and Pivovar [2] also reported the dynamic behavior of water in Nafion using quasi-elastic neutron scattering (QENS). The diffusion coefficents using QENS are about ten times greater than those reported by PFG-NMR. The reason for the difference between these methods is the time-dependent behavior of diffusion coefficients [3].

Proton transport and water diffusion in hydrated Nafion have also been examined by numerous molecular modeling studies [4–8]. In these papers, the residence time and diffusion coefficient of water in Nafion have been related to changes in the nanostructure of the water channel [9]. In addition, the dynamic properties of water, and the structure of the water channel with each hydration were in good agreement with experimental results.

Although Nafion is one candidate for practical use of PEM, sulfonated hydrocarbon polymers such as polysulfones [10], polyimides [11], and poly(ether ketone)s [12], etc. have been developed as alternatives to perfluorinated membranes due to their good properties. However, relatively few attempts have been made to understand the proton transport mechanism of these hydrocarbon polymers. Simpler sulfonated polyethersulfone (SPES) membranes serve as the starting point for our understanding of the proton transport mechanism in hydrocarbon systems.

The effect of hydration on the nanostructure of Nafion and SPES membranes has already been described in the first part of our work regarding the radial distribution function and cluster size distribution [13]. We examined the effect of hydration on membrane nanostructure and showed that water and hydronium ions in SPESs are bound strongly to the sufonate group for a lower λ compared with that of Nafion.

In the present work, we report data on the self-diffusion coefficient of water in both Nafion and SPES membranes using PFG-NMR experiments and molecular dynamics (MD) simulations. In particular, we discuss the differences between the dynamic properties of Nafion membranes and those of SPES in terms of both experimental results and classical MD simulations with an all-atom force field at various hydration levels. In addition, the mean residence times of H_2O and H_3O^+ were calculated from these simulations. Particular attention was given to the ion-exchange capacity (IEC) of SPES and water content. On the basis of these parameters, the dynamics of H_2O and H_3O^+ are discussed from the perspective of the self-diffusion coefficient and residence time.

Experimental and computational methodology

Commercially available perfluorinated Nafion membranes (NR-212) were purchased from DuPont (http://www2.dupont.com) and used as received. SPES of a hydrocarbon-type polymer membranes with different IEC values were obtained from Sumitomo Chemical (http://www.sumitomo-chem.co.jp/english/). These membranes were the same as those used in the first part of our paper.

The membrane was cut into thin 2.5×5 mm rectangular strips. Pieces of the cut membrane were immersed in a 3% H_2O_2 aqueous solution at 353 K for 1 h, and then rinsed with deionized water at 353 K for 1 h. Next the membranes were immersed in a 1.0 M H_2SO_4 aqueous solution for 1 h and rinsed with deionized water at 353 K for 1 h. After removing surface water by filter paper, eight cut pieces of the membrane are put into a glass tube (4 mm outer diameter, 40 mm long). To avoid the properties of the

membranes from being affected by the application of heat to the membrane, the membranes were not dried. Because SPES with IEC=1.4 meq/g cannot maintain the solid state in a 1.0 M H_2SO_4 aqueous solution, we report diffusion data for only two SPES membranes with IEC= 1.0 and IEC=0.75 meq/g.

To control the water content of the membranes, we used a humidity chamber (Tabai Espec, Osaka, Japan). Relative humidity (RH) was set at 50, 60, 70, 80, 90 and 95% at 30°C. The experimental diffusion data were displayed as a function of the number of water molecules per sulfonic acid group (λ). The λ value was calculated from the IEC and water content.

To estimate the ^1H self-diffusion coefficient of water, all measurements were performed with a narrow-bore NMR spectrometer (JEOL ECA-500) at 11.7 T equipped with a maximum gradient strength of 1,330 G/cm. The self-diffusion coefficients were measured with the bipolar pulse longitudinal eddy-current-delay pulse sequence with half-sine-shaped gradient pulses [3]. The diffusion time of the experiment was a pulse sequence of 20 ms.

We performed MD simulations with the LAMMPS code (http://lammps.sandia.gov) from Plimpton at Sandia [14]. The forcefield selected for this study is a simplified consistent forcefield (CFF), which is a second generation forcefield [15] similar to COMPASS [16]. The molecular modeling and forcefields parameter used were the same as those used in the first part of our paper [13]. The equations of motion were integrated using the Verlet algorithm [17] with a time step of 1.0 fs, and the particle-particle particle-mesh method [18] was employed to calculate electrostatic interactions. For each of the different systems, the MD simulation was performed at 313.15 K, which is the same temperature used in the water uptake and proton conductivity measurements. The simulations were performed as follows: the systems were first equilibrated for 1 ns with a step of 1 fs using the NVT ensemble (fixed volume and Nose-Hoover thermostat [19–22]). This was followed by a 15-ns MD simulation using the NPT ensemble (fixed pressure at 101,325 Pa and Nose-Hoover thermostat). The trajectories obtained from the last 5 ns of the NPT ensemble, measured every 5 ps were used for computing the dynamic properties.

Results and discussion

Self-diffusion coefficient

Although the results obtained from the MD simulation did not provide insight into the proton-hopping mechanism, investigation of water diffusion in the system was useful for understanding water behavior in PEMs. Protons are trans-

Fig. 1 a,b Mean square displacement (MSD) of $\lambda = 3$ plotted as log versus log time. The *red solid line* is the calculated region for the diffusion coefficient. **a** H_2O, **b** H_3O^+
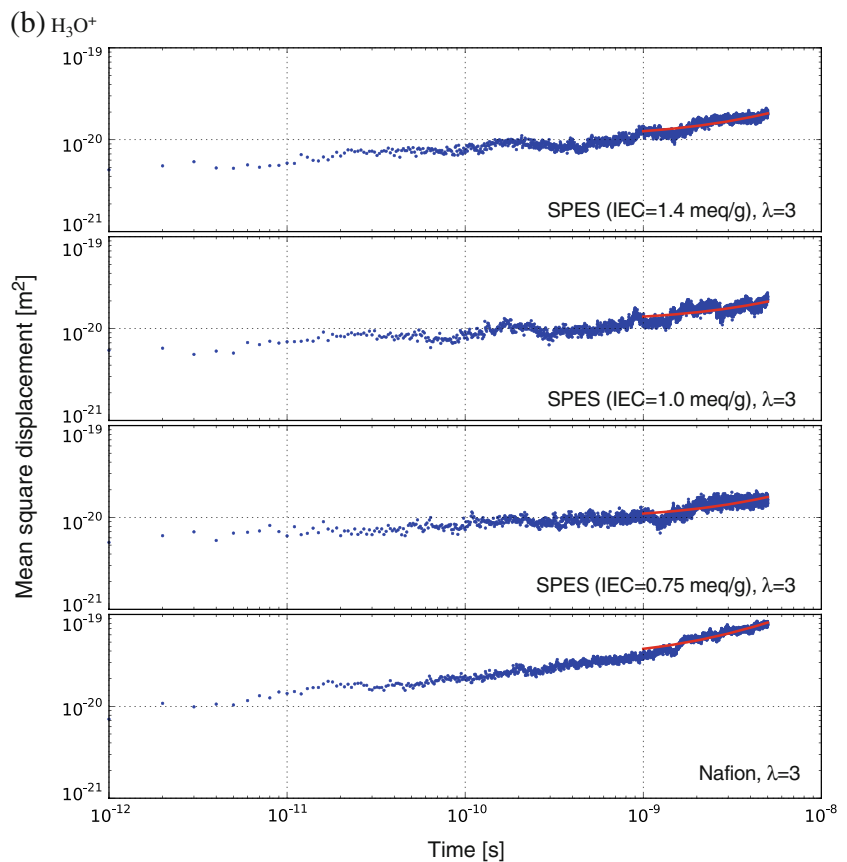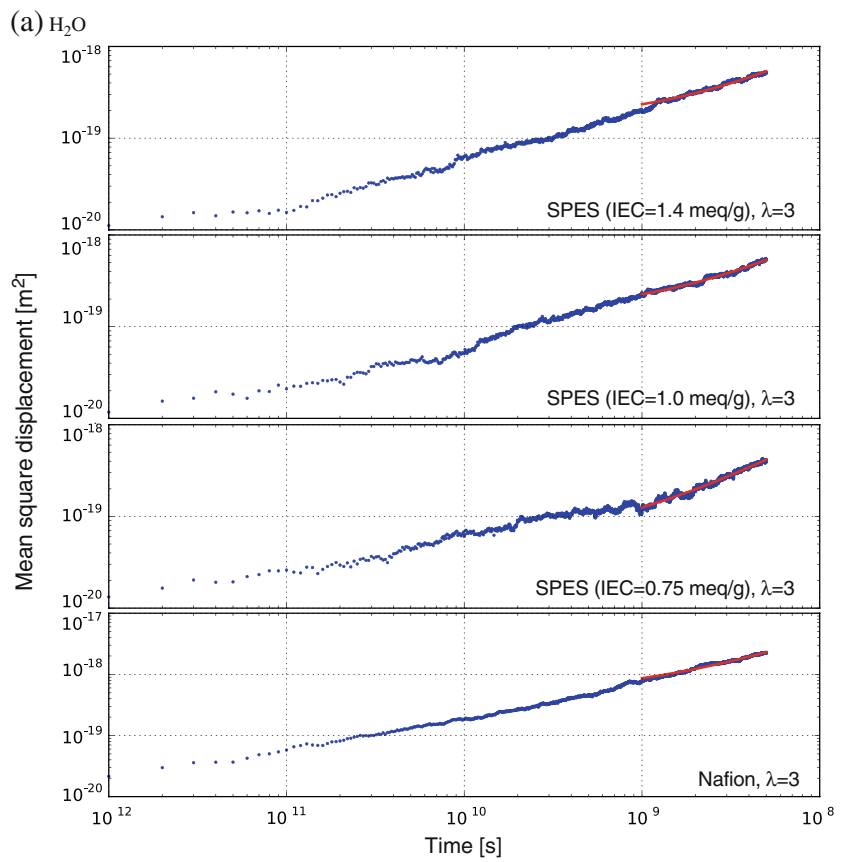

(a) H_2O

(b) H_3O^+

Table 1 Diffusion coefficients (D) of H_2O and H_3O^+ for Nafion and sulfonated polyethersulfone (SPES) membranes with different ion-exchange capacities (IEC) calculated from molecular dynamics (MD) simulations and experimental data (exp.) by NMR as a function of λ. Error was estimated from the standard error of the slope for linear fitting of mean square displacement (MSD). The uncertainty in a given value of diffusion coefficient by NMR is at least ~5%

λ	D(exp.)[10^{-9} m²/s]	D (H_2O) [10^{-9} m²/s]	D (H_3O^+) [10^{-9} m²/s]
Nafion			
2.38	0.0893	-	-
3	-	0.0601±0.0023	0.00165±0.00018
3.65	0.193	-	-
4.46	0.255	-	-
5.55	0.337	-	-
6	-	0.273±0.003	0.0198±0.0012
7.39	0.450	-	-
9	-	0.736±0.008	0.128±0.005
12	-	1.17±0.01	0.171±0.006
13.20	0.578	-	-
15	-	1.27±0.02	0.369±0.014
15.43	0.687	-	-
SPES (IEC=0.75 meq/g)			
3	-	0.0121±0.0004	0.000236±0.00006
3.15	0.00555	-	-
4.36	0.00772	-	-
5.24	0.0101	-	-
6	-	0.0307±0.0016	0.000392±0.000101
6.39	0.0149	-	-
8.06	0.0199	-	-
9	-	0.0815±0.0014	0.00102±0.00033
11.43	0.0256	-	-
12	-	0.166±0.003	0.00281±0.00037
12.40	0.0301	-	-
15	-	0.233±0.003	0.00476±0.00068
SPES (IEC=1.0 meq/g)			
2.76	0.00584	-	
3	-	0.0129±0.0005	0.000258±0.000074
4.14	0.00952	-	
5.15	0.0138	-	
6	-	0.0388±0.0011	0.000389±0.000127
6.38	0.0214	-	
9	-	0.120±0.00141	0.00226±0.00034
9.72	0.0340	-	
12	-	0.278±0.00253	0.00515±0.00099
13.1	0.0420	-	
13.5	0.0645	-	
15	-	0.455±0.006	0.00752±0.00114
SPES (IEC=1.4 meq/g)			
3	-	0.0126±0.0006	0.000290±0.000051
6	-	0.0673±0.0027	0.00132±0.00021
9	-	0.221±0.002	0.00778±0.00046
12	-	0.522±0.005	0.0236±0.0011
15	-	0.751±0.006	0.0557±0.0024

ported in the polymer by the displacement of water—the so-called vehicle mechanism, which is estimated from the diffusion coefficients based on the mean square displace-ment (MSD) values of H_2O and H_3O^+ against elapsed time. The mean square displacement of λ=3 is plotted as log versus log time (see Fig. 1 for an example). The diffusion

(a) Nafion

(b) SPES (IEC=0.75 meq/g)

(c) SPES (IEC=1.0 meq/g)

(d) SPES (IEC=1.4 meq/g)
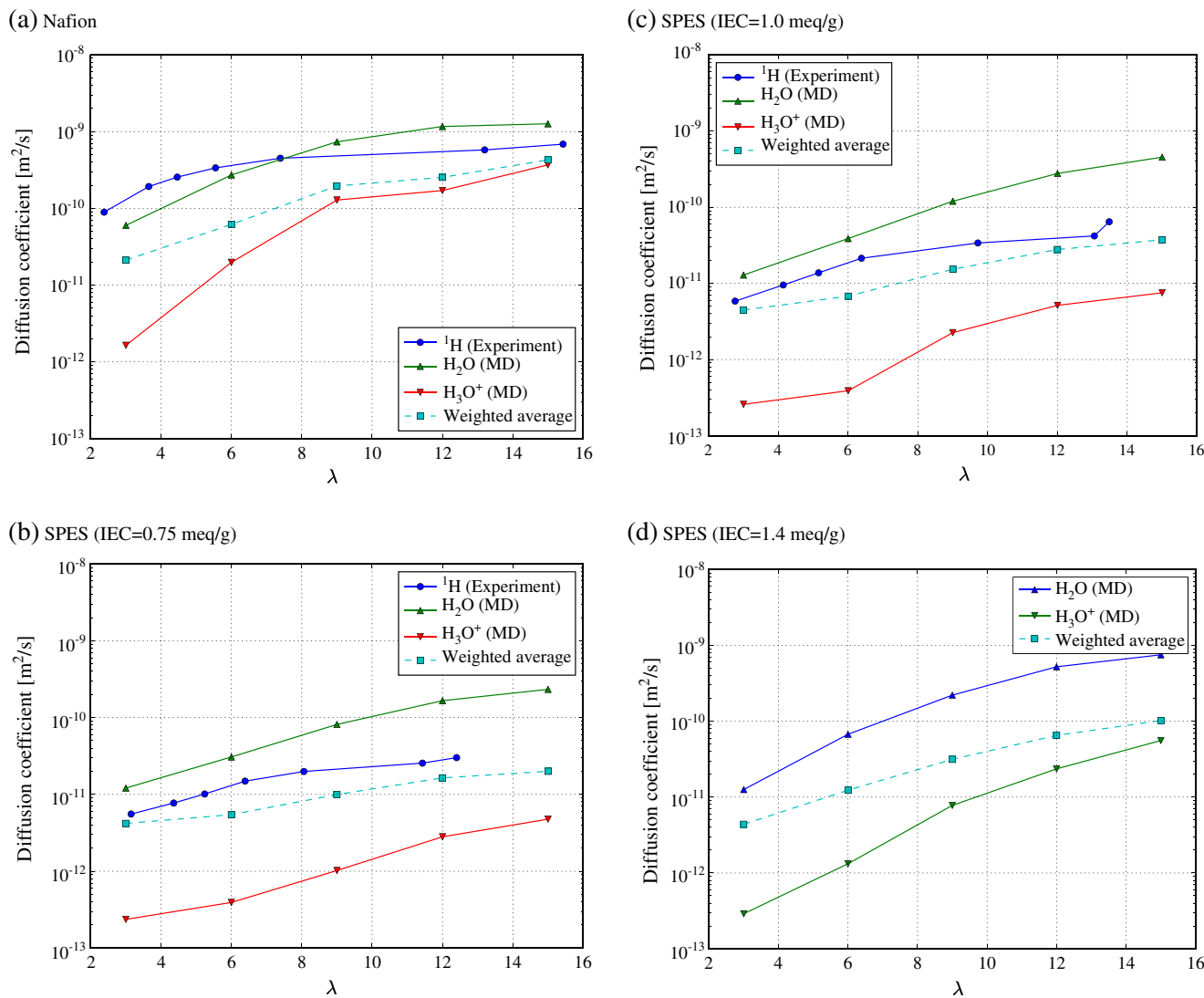
Fig. 2 Diffusion coefficients of H_2O and H_3O^+ calculated from MD simulations, the weighted average of H_3O^+ and H_2O in Eq. 1 and experimental data by NMR as a function of λ. **a** Nafion, **b** sulfonated polyethersulfone (SPES) with ion-exchange capacity (IEC)= 0.75 meq/g, **c** SPES with IEC=1.0 meq/g and **d** SPES with IEC= 1.4 meq/g

coefficients of H_2O and H_3O^+ are calculated from the linear region (red line in Fig. 1) of the MSD in elapsed time (3 ns at least) by using $\left\langle (r(t) - r(0))^2 \right\rangle = 6Dt$.

The diffusion coefficients estimated from the MD simulation and the PFG-NMR results are shown in Table 1 and Fig. 2 as a function of λ. For all membranes, the diffusion coefficient of H_3O^+ is about 5–20 times lower than that of H_2O. In particular, the diffusion for H_3O^+ compared with H_2O is even slower at lower λ. The electrostatic interaction of H_3O^+ with SO_3^- reduces the transport of H_3O^+ into the water channels, which was also observed from the structural properties. The diffusion coefficients of H_2O and H_3O^+ in Nafion have consistently higher values than those in the SPES for the same λ. For SPES, the solvent effect of H_2O for dissociating H_3O^+ from SO_3^- may be ineffective.

A clear difference was also observed in the diffusion coefficients at $\lambda=3$ and 6 for Nafion and SPES. The increase in hydration level from $\lambda=3$ to 6 corresponded to a 10-fold increase in the diffusion of Nafion, whereas the increase in diffusion was not as high for SPES, particularly for samples with a low IEC.

The cluster size distribution defined by the connectivity of water channels showed that most water in SPES was not incorporated into the large cluster to achieve better conductivity until the hydration level became sufficiently high [13]. The large cluster has the advantage of avoiding the tortuous structure of the diffusion path, i.e., water is displaced more easily in Nafion than in SPES. It is also expected that the formation of a large cluster in Nafion at $\lambda=6$ can cause the solvent effect that allows the efficient transport of H_3O^+ into the water channel, unlike in SPES.

(a) Nafion

(c) SPES (IEC=1.0 meq/g)
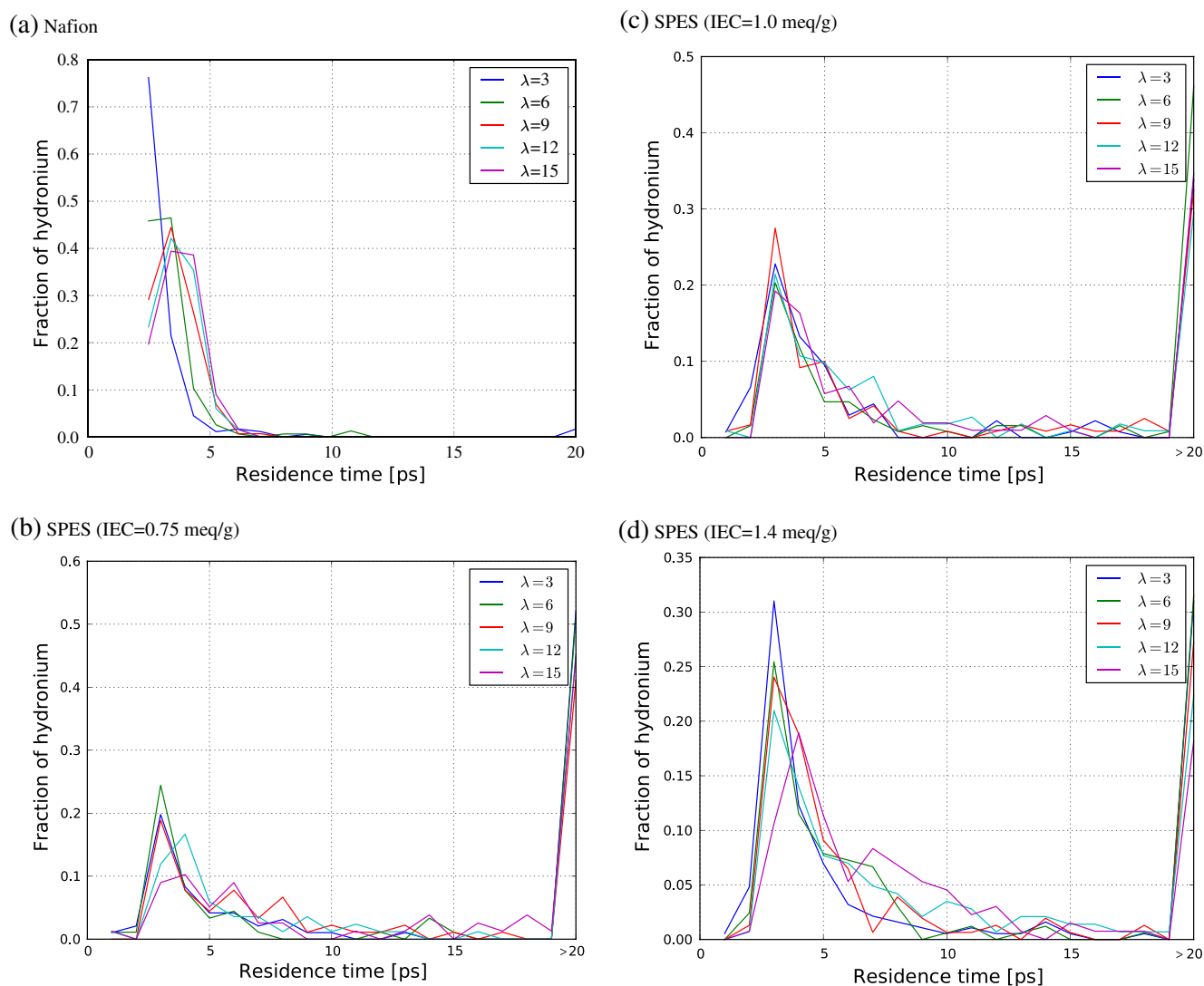

(b) SPES (IEC=0.75 meq/g)
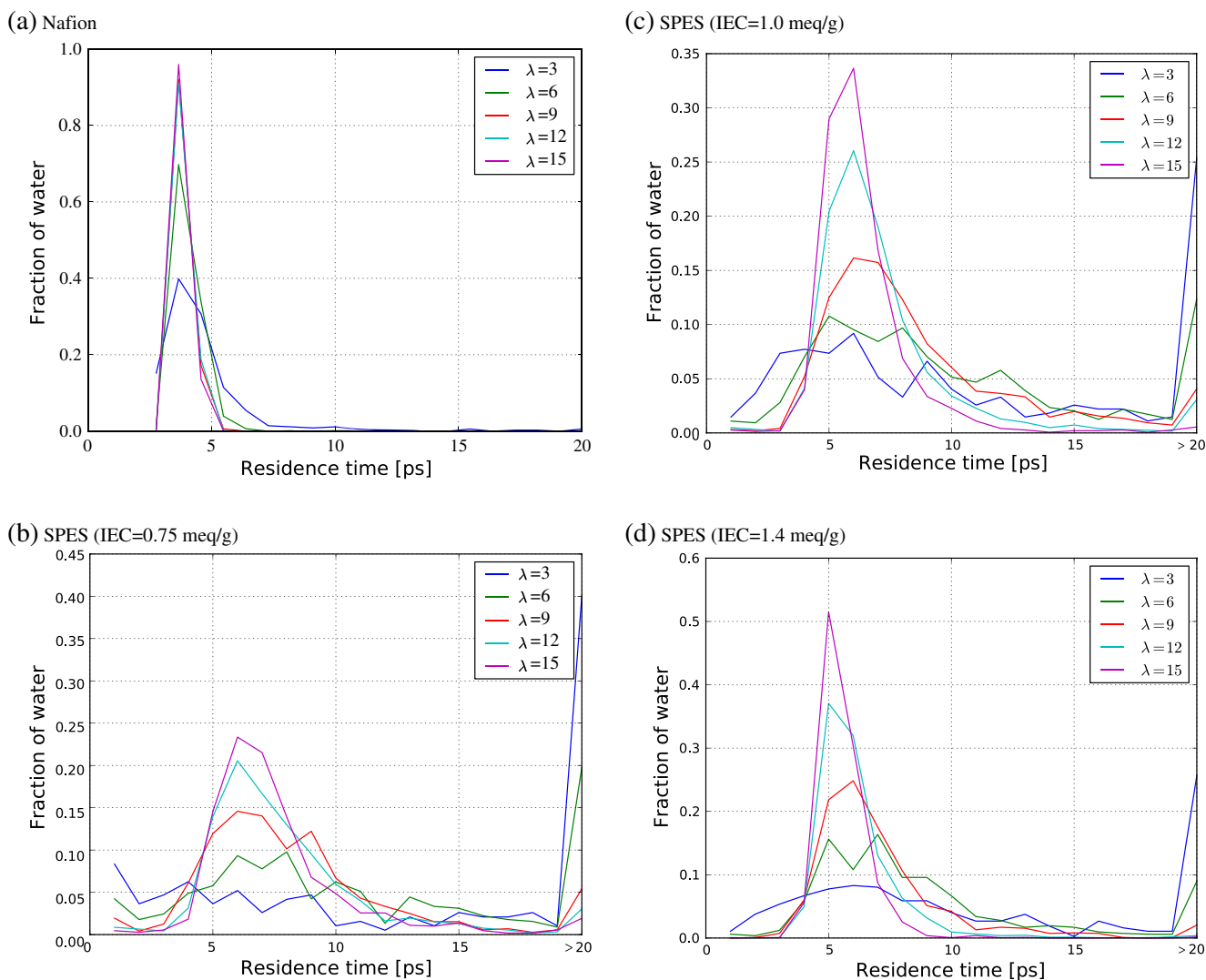
(d) SPES (IEC=1.4 meq/g)

Fig. 3 Distribution of residence times with 1 ps width of water in **a** Nafion, **b** SPES with IEC=0.75 meq/g, **c** SPES with IEC=1.0 meq/g ,and **d** SPES with IEC=1.4 meq/g at λ=3, 6, 9, 12 and 15. The integral of each residence time is normalized to 1

The diffusion coefficients of H_2O and H_3O^+ in Nafion have been calculated from MD simulations in the past. For example, Venkatnathan et al. and Devanathan et al. reported diffusion coefficients of H_2O and H_3O^+ at 300 and 350 K for a wide range of λ from 1 to 20. Cui et al. also reported the results for λ=3.4, 5.4, 8.6 and 12 at 300 K. Our values for the diffusion coefficient of H_2O are in very good agreement with these previous results, if the effect of temperature on diffusion is taken into consideration. The diffusion coefficients of H_3O^+ at λ=6 and 9 in the present study are close to those reported by Cui et al. [23], but our results differ from those of Venkatnathan et al [4]. This difference may be due to the hydronium model used in their calculations.

To understand the dynamic properties of protons, we can compare the proton diffusion coefficient obtained by PFG-NMR experiments with that obtained by MD simulations

for Nafion and SPES (IEC=0.75 and 1.0 meq/g). Experimental data is expected to deviate from H_3O^+ diffusion data at lower λ, which means that the H_3O^+ species cannot contribute to proton diffusion. For a system in fast exchange of protons between molecules, the observable proton diffusion coefficient, D_{exp}, by PFG-NMR can be described by a single average diffusion coefficient [24] due to the longer NMR timescale (millisecond-order), which is the weighted average of diffusion species such as H_3O^+ and H_2O. Weighting is done by using the molar fractions involved in a given molecule in each environment, so that,

$$D_{exp} = pD_{H_2O} + (1-p)D_{H_3O+} \tag{1}$$

where p is the fraction of H_2O in all molecules including exchangeable protons. Based on this interpretation, the results expected from MD calculations were displayed as the dotted red line in Fig. 2. The weighted average diffusion

(a) Nafion

(c) SPES (IEC=1.0 meq/g)

(b) SPES (IEC=0.75 meq/g)

(d) SPES (IEC=1.4 meq/g)

Fig. 4 Distribution of residence times with 1 ps width of hydronium in **a** Nafion, **b** SPES with IEC=0.75 meq/g, **c** SPES with IEC=1.0 meq/g , and **d** SPES with IEC=1.4 meq/g at λ=3, 6, 9, 12 and 15. The integral of each residence time is normalized to 1

coefficients were closer to the proton diffusion coefficients determined from PFG-NMR data, which were in good agreement with the experiment. The discrepancy between the experiment and MD simulation can be explained by the fact that PFG-NMR measurements probe diffusion behavior on the timescale of milliseconds rather than the nanoseconds used in MD simulations, and thus may be more sensitive to hindered motion between the water and polymer that occurs on a microscopic scale [3, 25].

Residence times of H_2O and H_3O^+

The mobility of H_2O and H_3O^+ were also analyzed by considering the distribution of their residence times around SO_3^-. We define residence time as the time spent by the diffusing species (H_2O and H_3O^+) within a certain cutoff distance from any SO_3^- group. The cutoff distance was set

to 4.8 Å based on the first minimum in the radial distribution functions of SO_3^- S-H_3O^+ and SO_3^-–H_2O shown in the first part of our paper [13].

The distribution of residence times of H_2O in Nafion and SPES with different IECs at λ=3, 6, 9, 12, and 15 are displayed in Fig. 3. In the figure, the curves are normalized to the total number of H_2O or H_3O^+ molecules, and residence times greater than 20 ps are shown at the far right of the figure. The residence time distribution of water in Nafion has the highest frequency at 3 ps for all λ; however, H_2O molecules with residence times greater than 20 ps were scarce in the case of Nafion. Distribution of the residence times at λ=3 in the range of 2–6 ps was observed for Nafion, which converges to 3 ps with increasing water content. In contrast, a broad distribution of H_2O residence times was observed in the range 1–15 ps for SPES, regardless of IEC. About 40% of the H_2O molecules for

SPES with IEC=0.75 meq/g and 25% of H_2O for SPES with IEC=1.0 and 1.4 meq/g had residence times greater than 20 ps. It is clear that water diffusion was suppressed as a result of the greater residence times around SO_3^-

The distribution of residence times of H_3O^+ is shown in Fig. 4. Residence times greater than 20 ps are plotted on the far right of the figures as in the case of H_2O. For Nafion, H_3O^+ with residence times greater than 20 ps was scarce at $\lambda=3$ and negligible at $\lambda=6$. In contrast, for SPES, there were many H_3O^+ molecules with longer residence times even when hydration levels were high. As with Nafion, a sufficient amount of water in SPES can cause dissociation of H_3O^+ from SO_3^-; however, a considerable number of H_3O^+ molecules had residence times greater than 20 ps, even at high levels of hydration.

A strong electrostatic interaction between H_3O^+ and SO_3^- was expected, considering the positive charge of H_3O^+ and negative charge of SO_3^-. At all hydration levels examined, the H_3O^+ molecule spends some time being (>20 ps) bond to the SO_3^- group, but this time decreases with increasing λ. The residence times can explain proton transport process reasonably, are consistent with the diffusion properties of H_2O and H_3O^+.

Conclusions

Differences in diffusion coefficients and residence times of water and hydronium were observed. With increasing IEC of SPES, some improvement in dynamic properties was achieved because of the formation of a well-organized water channel. It is likely that the location of SO_3^- on the flexible side chain of Nafion is more advantageous for efficient water diffusion compared with the location of SO_3^- on the main chain, as in the case of SPES. The weighted average diffusion coefficients of H_2O and H_3O^+ were closer to proton diffusion coefficients determined from the PFG-NMR data, which were in good agreement with the experiment.

MD simulations were useful for understanding the structural and dynamic properties on the molecular level. In future work, MD simulations based on this study will be used to investigate other advanced PEM such as side-chain-type hydrocarbon membranes.

References

1. Zawodzinski TA Jr, Neeman M, Sillerud LO, Gottesfeld S (1991) J Phys Chem B 95(15):6040–6047
2. Pivovar AA, Pivovar BS (2005) J Phys Chem B 109(2):785–793
3. Ohkubo T, Kidena K, Ohira A (2008) Macromolecules 41 (22):8688–8693
4. Devanathan R, Venkatnathan A, Dupuis M (2007) J Phys Chem B 111(45):13006–13013
5. Venkatnathan A, Devanathan R, Dupuis M (2007) J Phys Chem B 111(25):7234–7244
6. Jang SS, Molinero V, Cagin T, Goddard WA (2004) J Phys Chem B 108(10):3149–3157
7. Urata S, Irisawa J, Takada A, Shinoda W, Tsuzuki S, Mikami M (2005) J Phys Chem B 109(9):4269–4278
8. Cui S, Liu J, Selvan ME, Keffer DJ, Edwards BJ, Steele WV (2007) J Phys Chem B 111(9):2208–2218
9. Devanathan R, Venkatnathan A, Dupuis M (2007) J Phys Chem B 111(28):8069–8079
10. Zaidi SMJ, Mikhailenko SD, Robertson GP, Guiver MD, Kaliaguine S (2000) J Membr Sci 173(1):17–34
11. Gil M, Ji XL, Li XF, Na H, Hampsey JE, Lu YF (2004) J Membr Sci 234(1–2):75–81
12. Li L, Zhang J, Wang YX (2003) J Membr Sci 226(1–2):159–167
13. T Ohkubo, K Kidena, N Takimoto, Ohira A (2011) J Mol Model 17:739–755
14. Plimpton S (1995) J Comput Phys 117(1):1–19
15. Maple JR, Hwang MJ, Stockfisch TP, Dinur U, Waldman M, Ewig CS, Hagler AT (1994) J Comput Chem 15(2):162–182
16. Sun H (1998) J Phys Chem B 102(38):7338–7364
17. Verlet L (1967) Phys Rev 159:98–103
18. Hockney RW, Eastwood JW (1981) Computer simulation using particles. McGraw-Hill, New York
19. Nose S (1986) Mol Phys 57(1):187–191
20. Nose S (1984) J Chem Phys 81(1):511–519
21. Nose S (1984) Mol Phys 52(2):255–268
22. Nose S, Klein ML (1983) J Chem Phys 78(11):6928–6939
23. Cui S, Liu J, Selvan ME, Paddison SJ, Keffer DJ, Edwards BJ (2008) J Phys Chem B 112(42):13273–13284
24. Duval FP, Porion P, Faugere AM, Van Damme H (2001) J Colloid Interface Sci 242(2):319–326
25. Rollet AL, Simonin JP, Turq P, Gebel G, Kahn R, Vandais A, Noel JP, Malveau C, Canet D (2001) J Phys Chem B 105 (19):4503–4509

σ-Holes, π-holes and electrostatically-driven interactions

Jane S. Murray · Pat Lane · Timothy Clark ·
Kevin E. Riley · Peter Politzer

Abstract A positive π-hole is a region of positive electrostatic potential that is perpendicular to a portion of a molecular framework. It is the counterpart of a σ-hole, which is along the extension of a covalent bond to an atom. Both σ-holes and π-holes become more positive (a) in going from the lighter to the heavier atoms in a given Group of the periodic table, and (b) as the remainder of the molecule is more electron-withdrawing. Positive σ- and π-holes can interact in a highly directional manner with negative sites, e.g., the lone pairs of Lewis bases. In this work, the complexes of 13 π-hole-containing molecules with the nitrogen lone pairs of HCN and NH_3 have been characterized computationally using the MP2, M06-2X and B3PW91 procedures. While the electro-static interaction is a major driving force in π-hole bonding, a gradation is found from weakly noncovalent to considerably stronger with possible indications of some degree of coordinate covalency.

J. S. Murray (✉) · P. Politzer
CleveTheoComp, 1951 W,
59th Street Suite 409,
Cleveland, OH 44113, USA
e-mail: jsmurray@uno.edu

P. Lane
Department of Chemistry, University of New Orleans,
New Orleans, LA 70148, USA

T. Clark
Computer-Chemie-Centrum and Interdisciplinary Center for
Molecular Materials, Friedrich-Alexander-Universität,
Erlangen-Nürnberg,
Nägelsbachstraße 25,
91052 Erlangen, Germany

T. Clark
Centre for Molecular Design, University of Portsmouth,
Mercantile House,
Portsmouth PO1 2EG, UK

K. E. Riley
Department of Chemistry, University of Puerto Rico,
P.O. Box 23346, Rio Piedras, PR 00931, USA

σ-Holes

The term "σ-hole" was introduced by Clark et al. to describe the regions of positive electrostatic potential that are present on the outer surfaces of many covalently-bonded halogens [1]. These positive regions, discovered by Brinck et al. in 1992 [2], can interact electrostatically with negative sites on the same or more often other molecules, e.g., lone pairs and π-electrons, giving rise to noncovalent "halogen bonding" [3–5].

A σ-hole is formed when a halogen atom participates in a covalent sigma bond [5]. The accompanying rearrangement of the atom's electronic density typically leaves a region of diminished negative charge on its outer (non-involved) side. This region, which is along the extension of the bond, is called a σ-hole. Such charge anisotropy has indeed long been recognized [6–12]; when it is sufficient, a region of positive electrostatic potential results (a positive σ-hole).

Although the σ-hole concept was originally used to help explain the seeming anomaly of an electronegative halogen interacting attractively with a negative site [1, 3–5], it has since been found to be applicable as well to covalently-bonded atoms in Groups VI, V and IV [5, 13–17]. Interactions between these regions of positive electrostatic potential (positive σ-holes) and negative sites on the same or another molecular system are labeled σ-hole bonding;

halogen bonding is a subset of this. σ-Holes are normally concentrated along the extensions of the covalent bonds to an atom. This can be seen in Fig. 1 for $SeCl_2$; it has two positive σ-holes on the selenium and one on each chlorine, on the extensions of the Cl-Se and Se-Cl bonds, respectively. Accordingly, the resulting interactions tend to be highly directional. For a σ-hole-bonded complex R–Y–-Z, where Z is the negative site, the angle R-Y-Z is usually near 180° (barring secondary interactions).

Within any one of the Groups IV – VII, σ-holes become more positive and their interactions stronger (a) in going from the lighter to the heavier elements, as polarizability increases and electronegativity decreases, and also (b) as the remainder of the molecule becomes more electron-withdrawing [5, 16, 17]. Thus the σ-holes on the selenium in $SeCl_2$ are more positive than those on the sulfur in SCl_2, but less positive than the selenium ones in SeF_2. For a given negative site, σ-hole bonding interaction energies have been shown to correlate well with the magnitudes of the positive σ-hole electrostatic potentials [18–20]. It should be noted, however, that σ-holes need not always be positive; if the charge anisotropy is not sufficient, a negative σ-hole can result. This is found, for example, on the fluorine in H_3C-F. For more extensive discussions of σ-holes and their interactions, see Politzer et al. [5] and Murray et al. [17].

π-Holes

The suggestion has been made earlier that some molecules may exhibit positive "π-holes" [5]: regions of low electronic density that are perpendicular to portions of a molecular framework instead of being along the extensions of bonds (as are σ-holes). It was shown already some time ago that there are positive electrostatic potentials above the acyl carbons in H_3C-C(=O)F and H_3C-C(=O)NH_2 [21], which correlate with their relative tendencies to undergo

hydrolysis. Other candidates for positive π-holes might include SO_2 and SeO_2, if the π bonding electrons are drawn sufficiently toward the oxygens so that there are positive potentials above the sulfur and selenium. Indeed Fig. 2 shows that there is a positive electrostatic potential, a positive π-hole, above (and below) the selenium in SeO_2. Will this interact attractively with a negative site?

Our objective in this work has been to test for the presence of π-holes in a series of molecules of different types, and to investigate their interactions with the nitrogen lone pairs in HCN and NH_3. The first step is to compute the electrostatic potential on an appropriate outer surface of each molecule of interest.

The electrostatic potential

The nuclei and electrons of a molecule create an electrostatic potential V(r) in the surrounding space, given by Eq. 1:

$$V(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|} \qquad (1)$$

Z_A is the charge on nucleus A, located at $\mathbf{R}_A$, and $\rho(\mathbf{r})$ is the electronic density. V(r) is a physical observable, which can be determined experimentally by diffraction methods [22, 23] as well as computationally. The sign of V(r) in any region depends upon whether the positive effect of the nuclei or the negative one of the electrons is dominant there.

The electrostatic potential is an effective means for analyzing and predicting noncovalent interactions [24–26], which are largely electrostatically driven. For this purpose, V(r) is commonly computed on an appropriate outer surface of the molecule, since this is what will be seen by other approaching species. Following the suggestion of Bader et al. [27], the 0.001 au (elec-
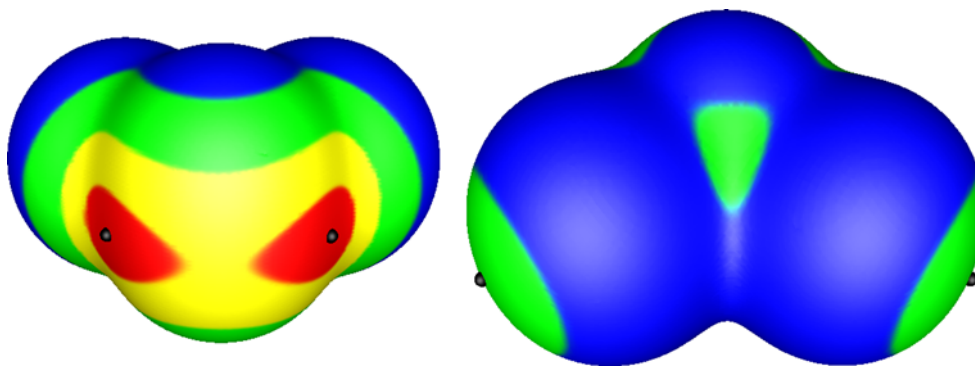
Fig. 1 Computed electrostatic potential on the 0.001 au molecular surface of $SeCl_2$. Two views are shown: (left) the selenium is in the foreground; (right) the chlorines are pointing to the left and right. Color ranges, in kcal mol^{-1}, are: red, greater than 26; yellow, from 26 to 13; green, from 13 to 0; blue, less than 0 (negative). The positions of the most positive electrostatic potentials associated with the σ-holes on the selenium and the chlorines are shown as black hemispheres. $V_{S,max}$ (Se)=32.5 kcal mol^{-1}; $V_{S,max}$ (Cl)=10.4 kcal mol^{-1}
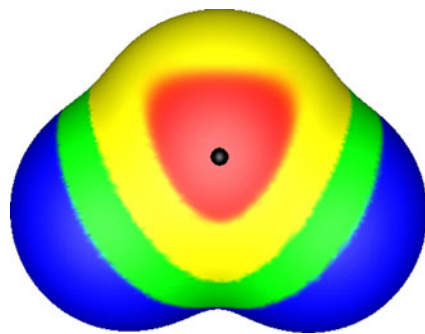
Fig. 2 Computed electrostatic potential on the 0.001 au molecular surface of SeO_2. The selenium is in the middle. Color ranges, in kcal mol^{-1}, are: red, greater than 33; yellow, from 33 to 20; green, from 20 to 0; blue, less than 0 (negative). The position of the most positive electrostatic potential associated with the π-hole above and below the selenium is indicated by a black hemisphere; $V_{S,max}$=35.4 kcal mol^{-1}

trons/bohr^3) contour of the molecule's electronic density is frequently taken to be the surface. This has the advantage of being specific to the particular molecule, and thus reflecting features such as lone pairs, π electrons and strained bonds. The 0.001 au contour normally lies beyond the van der Waals radii of the main group atoms [28] (except for hydrogen), which makes it suitable for noncovalent interactions.

Procedure

We have computed the electrostatic potentials on the 0.001 au surfaces of the molecules listed in Table 1. The B3PW91/6-31 G(d,p) procedure was used, to allow comparisons with earlier calculated σ-hole potentials [4, 5, 13–16, 18, 28]. The wavefunctions were obtained with Gaussian 09 [29] and the surface potentials, labeled $V_S(\mathbf{r})$, with the Wave Function Analysis-Surface Analysis Suite [30]. The latter code gives the magnitudes and positions of the locally most positive and most negative values of $V_S(r)$, designated $V_{S,max}$ and $V_{S,min}$; there can be more than one of each on any given molecular surface.

The geometries and interaction energies ΔE of the complexes formed between the molecules in Table 1 and the Lewis bases HCN and NH_3 were determined by three different computational techniques: MP2, M06-2X [31] and B3PW91, in conjunction with the aug-cc-pVDZ basis set. The interaction energies were obtained from the molecular energy minima at 0 K with Eq. 2:

$$\Delta E = E(complex) - E(\pi - hole\ molecule)$$
$$- E(Lewis\ base) \quad (2)$$

In view of the large basis set being used, basis set superposition error should be minimal [32] and was accordingly not considered.

Results

Electrostatic potentials

For each of the molecules in Table 1 except for H_2CO, regions of positive electrostatic potential with $V_{S,max}$ are found above and below the central atom; these are what we label positive π-holes. (In H_3CPO_2 and $H_5C_6PO_2$, the phosphorus is considered to be the central atom.) The π-holes of SeO_2 and H_3CPO_2 are displayed in Figs. 2 and 3.

The $V_{S,max}$ associated with the π-holes are given in Table 1. Their magnitudes are comparable to, and in some cases considerably exceed, those of σ-holes [4, 5, 13–16, 18]. As with the latter, π-hole $V_{S,max}$ become more positive in going from the lighter to the heavier atoms in a given column of the periodic table (compare O_3, SO_2 and SeO_2). Note that H_2CO does not even have a π-hole, whereas H_2SiO has quite a strong one. The $V_{S,max}$ also become more positive as the remainder of the molecule is more electron-withdrawing (compare F_2CO and Cl_2CO). The sizable difference between the π-hole $V_{S,max}$ of H_3CPO_2 and $H_5C_6PO_2$ is probably due to overlapping of the positive π-hole potential and the positive potentials of the nearby methyl hydrogens in H_3CPO_2.

Complexes with HCN and NH_3

In Tables 2 and 3 are some key properties of the complexes formed by the molecules in Table 1 (except H_2CO) with the nitrogen lone pairs of HCN and NH_3, as computed by the different procedures mentioned above. The MP2 and the

Table 1 Computed electrostatic potential maxima ($V_{S,max}$) on 0.001 au molecular surfaces,[a] above and below indicated atom (π-holes)

Molecule	Atom with π-hole	$V_{S,max}$ (kcal/mol)
H_2CO	C	b
Cl_2CO	C	22.8 (2)
F_2CO	C	40.9 (2)
H_2SiO	Si	43.4 (2)
F_2SiO	Si	66.7 (2)
FNO_2	N	32.8 (2)
$H_5C_6PO_2$	P	30.2 (2)
H_3CPO_2	P	48.6, 46.5
FPO_2	P	58.4 (2)
O_3	central O	22.9 (2)
SO_2	S	32.9 (2)
SeO_2	Se	35.4 (2)

[a] B3PW91/6-31 G(d,p)

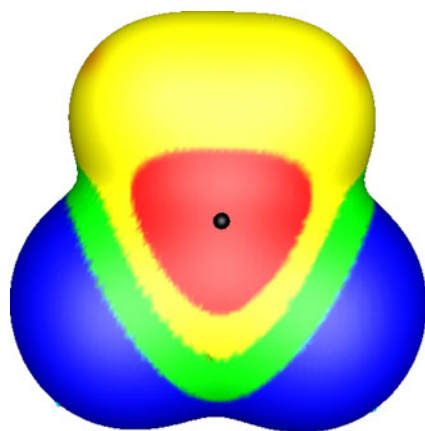[b] The surface above and below the carbon is positive but there is no $V_{S,max}$ present

Fig. 3 Computed electrostatic potential on the 0.001 au molecular surface of H_3CPO_2. The phosphorus is in the middle. Color ranges, in kcal mol^{-1}, are: red, greater than 33; yellow, from 33 to 20; green, from 20 to 0; blue, less than 0 (negative). The position of the most positive electrostatic potential associated with the π-hole on the side of the phosphorus that is shown is indicated by a black hemisphere; $V_{S,max}=46.5$ kcal mol^{-1}

M06-2X usually predict somewhat stronger binding (more negative ΔE) than does the B3PW91.

The MP2 method is known to give reasonably accurate binding energies and interaction geometries for a wide variety of complex types when used along with a medium sized basis set, such as cc-pVTZ or aug-cc-pVDZ [34, 35]. Thus, results obtained using the MP2/aug-cc-pVDZ method should be seen as the reference values in the current study, although these data cannot be said to be of benchmark quality. Most DFT methods, including B3PW91, are considered to not properly describe dispersion interactions within molecular complexes [36, 37]. The M06-2X functional, however, does account for dispersion at least semi-quantitatively and has been found to be effective for studies involving noncovalent interactions [31], although with a tendency for overbinding in some instances [37]. The most likely explanation for the underbinding seen for the B3PW91/aug-cc-pVDZ method in Tables 2 and 3 (as well as Table 4) is its inadequate treatment of dispersion, which certainly plays some role in stabilizing these complexes, even though the primary driving forces for these interactions, as will be demonstrated, are electrostatic.

The interaction of a given molecule with NH_3 is invariably more stabilizing than with HCN; this is evident in the more negative ΔE and the usually shorter separations observed with NH_3. One reason for this is that the electric field of the nitrogen lone pair in NH_3 is considerably stronger and more polarizing than that in HCN [38], which enhances the electrostatic interactions of the former.

Comparison of the $V_{S,max}$ in Table 1 with the ΔE in Tables 2 and 3 shows that, within Groups IV, V and VI taken separately, the interaction energies with both HCN and NH_3 become generally more negative (more stabilizing)

as the π-hole $V_{S,max}$ increases. This reflects the importance of the electrostatic factor in π-hole interactions.

The ΔE in Tables 2 and 3 cover a remarkably large range. While many are between 0 and -12 kcal mol^{-1}, as is expected for noncovalent interactions, some – especially among the complexes with NH_3 – are considerably more negative, with values of -22 to -46 kcal mol^{-1}. These surprisingly strong interactions are reflected in the ratios of the separations to the sums of the van der Waals radii of the respective atoms, which are as low as 0.5 to 0.6. It seems evident that some of these complexes are beyond what is expected of noncovalent interactions.

This conclusion is buttressed by their relatively large interaction angles (Tables 2 and 3). Since the molecules in Table 1 are all planar (except for the methyl hydrogens in H_3CPO_2), the π-hole concept implies that these angles should be about 90°, and this is indeed true of some of them. Deviations of a few degrees can be attributed to secondary interactions. For example, the O-C-N angle of about 97° in $F_2C(=O)$—NH_3 reflects some degree of repulsion between the oxygen and the nitrogen lone pair. Analogous explanations can be given in other instances of positive deviations from 90°. In O_2O—NCH and O_2O—NH_3, the MP2 and M06-2X interaction angles are significantly less than 90°; this can be rationalized by noting that the O–N separations predicted by these methods are small enough to allow the outer oxygens in O_3 to interact attractively with the carbon in HCN and one of the hydrogens in NH_3. The B3PW91 O–N separations are much greater, and the O–C and O–H interactions consequently appear to not be important.

While secondary interactions can explain many of the observed angular deviations from 90°, the larger ones may reflect an additional factor. A number of the complexes have interaction angles in the immediate vicinity of 100° or even larger, as much as 108°. These tend to be the same systems that have the most negative ΔE and the low values of the ratio of separation to sum of van der Waals radii, mentioned above. In each instance, the atom having the π-hole is a silicon or a phosphorus.

These complexes may have some degree of coordinate covalent character, with the HCN or NH_3 nitrogen sharing its lone pair to some extent. The atom that has the π-hole changes from a trigonal to a quasi-tetrahedral configuration. A coordinate covalent contribution to the interaction would be favored for NH_3 over HCN; the average local ionization energy [39] of the nitrogen lone pair in NH_3 is 7.3 eV at the B3PW91/6-31 G(d,p) level, compared to 10.7 eV for HCN. Accordingly the NH_3 lone pair electrons are more readily available. Thus the stronger interactions found for NH_3 can perhaps be explained in terms of both electrostatic (see above) and coordinate covalent considerations.

Table 2 Computed properties of π-hole complexes with HCN, using the MP2, M06-2X and B3PW91 methods and the aug-cc-pVDZ basis set

Complex	Interaction energy (kcal/mol)	Interaction separation (Å)	Ratio of interaction separation to sum of vdW radii[a]	Interaction angle (degrees)
Group IV				
Cl₂CO—NCH				
MP2	-3.99	C—N: 2.89	0.89	<O-C-N: 91.4
M06-2X	-2.99	C—N: 2.95	0.91	<O-C-N: 93.8
B3PW91	-0.50	C—N: 3.41	1.05	<O-C-N: 90.9
F₂CO—NCH				
MP2	-4.58	C—N: 2.77	0.85	<O-C-N: 97.4
M06-2X	-4.09	C—N: 2.71	0.83	<O-C-N: 99.1
B3PW91	-1.68	C—N: 3.01	0.93	<O-C-N: 98.0
H₂SiO—NCH				
MP2	-11.55	Si—N: 2.18	0.60	<O-Si-N: 103.6
M06-2X	-12.99	Si—N: 2.11	0.58	<O-Si-N: 103.0
B3PW91	-10.91	Si—N: 2.10	0.58	<O-Si-N: 102.7
F₂SiO—NCH				
MP2	-22.58	Si—N: 2.01	0.55	<O-Si-N: 107.4
M06-2X	-24.96	Si—N: 1.98	0.54	<O-Si-N: 105.5
B3PW91	-21.71	Si—N: 1.96	0.54	<O-Si-N: 104.9
Group V				
FNO₂—NCH				
MP2	-4.38	N—N: 2.80	0.90	<O-N-N: 89.2
M06-2X	-3.64	N—N: 2.77	0.89	<O-N-N: 91.4
B3PW91	-1.30	N—N: 3.12	1.01	<O-N-N: 88.9
H₅C₆PO₂—NCH				
MP2[b]	-5.42	P—N: 2.74	0.82	<O-P-N: 93.5
M06-2X	-5.37	P—N: 2.71	0.81	<O-P-N: 92.6
B3PW91	-2.30	P—N: 2.74	0.82	<O-P-N: 91.9
H₃CPO₂—NCH				
MP2	-7.59	P—N: 2.68	0.80	<O-P-N: 94.2
M06-2X	-7.04	P—N: 2.69	0.80	<O-P-N: 94.3
B3PW91	-4.04	P—N: 2.67	0.80	<O-P-N: 94.3
FPO₂—NCH				
MP2	-11.88	P—N: 2.21	0.66	<O-P-N: 97.8
M06-2X	-14.07	P—N: 2.07	0.62	<O-P-N: 99.0
B3PW91	-10.61	P—N: 2.09	0.62	<O-P-N: 98.8
Group VI				
O₂O—NCH[c]				
MP2	-3.04	O—N: 2.89	0.94	<O-O-N: 77.5
M06-2X	-2.47	O—N: 2.89	0.94	<O-O-N: 83.
B3PW91	-0.83	O—N: 3.32	1.08	<O-O-N: 99.2
O₂S—NCH				
MP2	-4.04	S—N: 3.03	0.90	<O-S-N: 94.1
M06-2X	-3.99	S—N: 2.92	0.87	<O-S-N: 94.2
B3PW91	-2.30	S—N: 3.11	0.93	<O-S-N: 99.0
O₂Se—NCH				
MP2	-4.14	Se—N: 3.08	0.89	<O-Se-N: 96.4
M06-2X	-5.27	Se—N: 2.94	0.85	<O-Se-N: 95.0
B3PW91	-3.36	Se—N: 3.05	0.88	<O-Se-N: 99.2

[a] The van der Waals radii are from reference [33]

[b] Computed with the cc-pVDZ basis set because of space limitations

[c] The central oxygen is interacting with the NCH

Table 3 Computed properties of π-hole complexes with NH_3, using the MP2, M06-2X and B3PW91 methods and the aug-cc-pVDZ basis set

Complex	Interaction energy (kcal/mol)	Interaction separation (Å)	Ratio of interaction separation to sum of vdW radii[a]	Interaction angle (degrees)
Group IV				
Cl_2CO—NH_3				
MP2	-4.64	C—N: 2.85	0.88	<O-C-N: 87.5
M06-2X	-4.44	C—N: 2.87	0.88	<O-C-N: 94.3
B3PW91	-1.18	C—N: 3.13	0.96	<O-C-N: 89.1
F_2CO—NH_3				
MP2	-6.14	C—N: 2.67	0.82	<O-C-N: 96.6
M06-2X	-6.57	C—N: 2.61	0.80	<O-C-N: 97.8
B3PW91	-3.33	C—N: 2.72	0.84	<O-C-N: 97.1
H_2SiO—NH_3				
MP2	-26.69	Si—N: 2.03	0.56	<O-Si-N: 102.7
M06-2X	-30.27	Si—N: 2.00	0.55	<O-Si-N: 101.3
B3PW91	-26.43	Si—N: 2.02	0.55	<O-Si-N: 102.6
F_2SiO—NH_3				
MP2	-42.38	Si—N: 1.95	0.53	<O-Si-N: 108.6
M06-2X	-46.19	Si—N: 1.93	0.53	<O-Si-N: 107.9
B3PW91	-41.17	Si—N: 1.94	0.53	<O-Si-N: 107.9
Group V				
MP2	-5.52	N—N: 2.79	0.90	<O-N-N: 93.5
M06-2X	-5.17	N—N: 2.77	0.89	<O-N-N: 92.5
B3PW91	-2.16	N—N: 3.01	0.97	<O-N-N: 91.2
$H_5C_6PO_2$—NH_3				
MP2 [b]	-17.63	P—N: 2.10	0.63	<O-P-N: 98.1
M06-2X	-21.61	P—N: 2.00	0.60	<O-P-N: 98.6
B3PW91	-15.96	P—N: 2.04	0.61	<O-P-N: 98.0
H_3CPO_2—NH_3				
MP2	-18.41	P—N: 2.05	0.61	<O-P-N: 98.2
M06-2X	-22.17	P—N: 2.00	0.60	<O-P-N: 98.3
B3PW91	-17.26	P—N: 2.04	0.61	<O-P-N: 98.2
FPO_2—NH_3				
MP2	-32.35	P—N: 1.95	0.58	<O-P-N: 101.7
M06-2X	-37.45	P—N: 1.92	0.57	<O-P-N: 101.7
B3PW91	-32.08	P—N: 1.94	0.58	<O-P-N: 101.7
Group VI				
O_2O—NH_3[c]				
MP2	-3.86	O—N: 2.95	0.96	<O-O-N: 72.4
M06-2X	-3.83	O—N: 2.94	0.96	<O-O-N: 74.3
B3PW91	-1.12	O—N: 3.23	1.05	<O-O-N: 99.1
O_2S—NH_3				
MP2	-6.00	S—N: 2.83	0.84	<O-S-N: 93.1
M06-2X	-8.11	S—N: 2.59	0.77	<O-S-N: 93.9
B3PW91	-6.42	S—N: 2.59	0.77	<O-S-N: 95.7
O_2Se—NH_3				
MP2	-7.62	Se—N: 2.87	0.83	<O-Se-N: 91.8
M06-2X	-12.79	Se—N: 2.40	0.70	<O-Se-N: 92.0
B3PW91	-10.12	Se—N: 2.51	0.73	<O-Se-N: 93.8

[a] The van der Waals radii are from reference [33]

[b] Computed with the cc-pVDZ basis set because of space limitations

[c] The central oxygen is interacting with the NH_3

Table 4 Computed properties of π-hole complexes of BF_3 and BCl_3 with HCN and NH_3, using the MP2, M06-2X and B3PW91 methods and the aug-cc-pVDZ basis set

Complex	Interaction energy (kcal/mol)	Interaction separation (Å)	Ratio of interaction separation to sum of vdW radii[a]	Interaction angle (degrees)
F_3B—NCH				
MP2	-6.94	B—N: 2.35	0.66	<F-B-N: 93.8
M06-2X	-6.64	B—N: 2.35	0.66	<F-B-N: 97.0
B3PW91	-3.61	B—N: 2.36	0.66	<F-B-N: 97.7
Cl_3B—NCH				
MP2	-4.46	B—N: 2.84	0.80	<Cl-B-N: 91.7
M06-2X	-3.52	B—N: 2.88	0.81	<Cl-B-N: 91.3
B3PW91	-0.64	B—N: 3.34	0.94	<Cl-B-N: 90.8
F_3B—NH_3				
MP2	-26.38	B-N: 1.65	0.46	<F-B-N: 104.3
M06-2X	-27.59	B-N: 1.65	0.46	<F-B-N: 104.3
B3PW91	-23.73	B-N: 1.65	0.46	<F-B-N: 104.4
Cl_3B—NH_3				
MP2	-29.67	B-N: 1.62	0.46	<Cl-B-N: 104.9
M06-2X	-31.29	B-N: 1.61	0.45	<Cl-B-N: 105.1
B3PW91	-25.04	B-N: 1.62	0.46	<Cl-B-N: 105.2

[a] The van der Waals radii are from reference [33]

Gradations in π-hole and σ-hole bonding

What we see in Tables 2 and 3 is that there is a gradation in the interactions between molecules having π-holes and Lewis bases. They can range from quite weak, with separations near the sums of the respective van der Waals radii and little or no effect upon the π-hole molecule's geometry, to relatively strong with small separations and significant distortion.

What determines where a given complex will fit into this spectrum of interactions? The results in Tables 2 and 3 indicate that the stronger interactions are more likely when the π-hole is on a relatively large atom, e.g., silicon or phosphorus. Such atoms are more polarizable and less electronegative, both of which lead to a larger initial $V_{S,max}$; high polarizability also increases the response of the atom to the electric field of the Lewis base. In addition, a second-row or larger atom can better accommodate a close approach of the base in a possible increase in coordination; in orbital terminology, the atom can become hypervalent. Finally, the chance of some degree of coordinate covalency is enhanced by the lone pair electrons of the base having low ionization energies and by the π-hole molecule having the capacity to partially share them, which is related to its overall polarizability [40].

Analogous considerations are applicable to interactions with σ-holes as well as with π-holes. We have seen examples in earlier work involving σ-holes on covalently-bonded silicon, for instance in the formation of complexes

between SiF_4 and amines [15, 41], and in an intramolecular rearrangement of a silicon nitrate ester [42]. Note also some of the silyl halide/NH_3 and silyl halide/H_2O systems studied by Ignatyev and Schaefer [43]. Recently Del Bene et al. have found a very marked gradation of halogen (σ-hole) bond strengths in a series of systems of the type FCl–CNX, where X represents various atoms or groups selected to give CNX a range of basicities [44].

An interesting example of gradations in bonding is provided by the complexes of BF_3 and BCl_3 with HCN and NH_3. BF_3 and BCl_3 clearly have positive π-holes, corresponding to the unoccupied boron 2pπ orbitals. The B3PW91/6-31 G(d,p) $V_{S,max}$ are 48.8 kcal mol^{-1} for BF_3 and 24.0 kcal mol^{-1} for BCl_3; the difference reflects the greater electron-withdrawing power of the fluorines.

Table 4 shows that there is quite a contrast between the complexes that BF_3 and BCl_3 form with HCN and those with NH_3. With HCN, the interactions are fairly weak; that of BF_3 is stronger, with ΔE about −7 kcal mol^{-1} vs. -4 kcal mol^{-1} for BCl_3, due to the much more positive $V_{S,max}$ of BF_3.

With NH_3, on the other hand, the complexes of both BF_3 and BCl_3 are as strongly bound as some of those involving silicon and phosphorus derivatives in Table 3. They have ΔE between −27 and −31 kcal mol^{-1}, short separations and interaction angles of 104° - 105°. This may seem inconsistent with the earlier statement that strong interactions are more likely with larger, more polarizable atoms, but it should be noted that boron is near the beginning of

the first row, and thus significantly larger and more polarizable than the atoms that follow it. (The polarizability of boron, 3.03 $Å^3$, is closer to that of phosphorus, 3.63 $Å^3$, than to that of carbon, 1.76 $Å^3$ [45].)

The greater stability of the NH_3 complexes compared to the HCN can be explained by the stronger electric field and lower ionization energy of the NH_3 lone pair electrons. What is notable, however, is that ΔE is slightly more negative for Cl_3B—NH_3 than for F_3B—NH_3. Since BF_3 has a much more positive $V_{S,max}$ than does BCl_3, it appears that some additional factor is involved. Brinck et al. suggested that this other factor is the higher "charge capacity" of BCl_3 [46]; by virtue of its larger polarizability (9.38 $Å^3$ $vs.$ 3.31 $Å^3$ [45]), BCl_3 is better able to accept some share of the NH_3 lone pair.

Conclusions

The concept of σ-holes has been extended by demonstrating the existence of π-holes, and showing that they have analogous properties. The interactions of a series of molecules having π-holes with the nitrogen lone pairs of HCN and NH_3 have been analyzed and described computationally.

A major driving force in both σ-hole and π-hole bonding is the electrostatic interaction between the positive σ- or π-hole and the negative site. However both types of bonds can show a significant gradation, ranging from weak, noncovalent and largely electrostatic to considerably stronger with evidence of some coordinate covalent character.

References

1. Clark T, Henneman M, Murray JS, Politzer P (2007) J Mol Model 13:291–296
2. Brinck T, Murray JS, Politzer P (1992) Int J Quantum Chem 44:55–64
3. Auffinger P, Hays FA, Westhof E, Ho PS (2004) Proc Natl Acad Sci USA 101:16789–16794
4. Politzer P, Lane P, Concha MC, Ma Y, Murray JS (2007) J Mol Model 13:305–311
5. Politzer P, Murray JS, Clark T (2010) Phys Chem Chem Phys 12:7748–7757
6. Stevens ED (1979) Mol Phys 37:27–45
7. Nyburg SC, Wong-Ng W (1979) Proc R Soc Lond A 367:29–45
8. Ikuta S (1990) J Mol Struct THEOCHEM 205:191–201
9. Price SL, Stone AJ, Lucas J, Rowland RS, Thornley AE (1994) J Am Chem Soc 116:4910–4918
10. Tsirelson VG, Zou PF, Tang T-H, Bader RWF (1995) Acta Crystallogr A 51:143–153
11. Lommerse JPM, Stone AJ, Taylor R, Allen FH (1996) J Am Chem Soc 118:3108–3116
12. Grabowski SJ, Bilewicz E (2006) Chem Phys Lett 427:51–55
13. Murray JS, Lane P, Clark T, Politzer P (2007) J Mol Model 13:1033–1038
14. Murray JS, Lane P, Politzer P (2007) Int J Quantum Chem 107:2286–2292
15. Murray JS, Lane P, Politzer P (2009) J Mol Model 15:723–729
16. Politzer P, Murray JS (2009) In: Leszczynski J, Shukla M (eds) Practical Aspects of Computational Chemistry. Springer, Heidelberg, pp 149–163
17. Murray JS, Riley KE, Politzer P, Clark T (2010) Aust J Chem 63:1598–1607
18. Riley KE, Murray JS, Concha MC, Politzer P, Hobza P (2009) J Chem Theor Comput 5:155–163
19. Shields ZP, Murray JS, Politzer P (2010) Int J Quantum Chem 110:2823–2832
20. Riley KE, Murray JS, Fanfrlík J, Řezáč J, Solá RJ, Concha MC, Ramos FM, Politzer P (2011) J Mol Model, doi:10.1007/s00894-011-1015-6
21. Sjoberg P, Politzer P (1990) J Phys Chem 94:3959–3961
22. Stewart RF (1979) Chem Phys Lett 65:335–342
23. Politzer P, Truhlar DG (eds) (1981) Chemical Applications of Atomic and Molecular Electrostatic Potentials. Plenum Press, New York
24. Naray-Szabo G, Ferenczy GG (1995) Chem Rev 95:829–847
25. Politzer P, Murray JS (2002) Theor Chem Acc 108:134–142
26. Murray JS, Politzer P (2011) Rev Comput Mol Sci 1:153–163
27. Bader RFW, Carroll MT, Cheeseman JR, Chang C (1987) J Am Chem Soc 109:7968–7979
28. Murray JS, Politzer P (2009) Croat Chem Acta 82:267–275
29. Frisch MJ et al. (2009) Gaussian 09. Gaussian Inc, Wallingford, CT
30. Bulat FA, Toro-Labbé A, Brinck T, Murray JS, Politzer P (2010) J Mol Model 16:1679–1691
31. Zhao Y, Truhlar DG (2008) Theor Chem Acc 120:215–241
32. Grimme S (2006) J Comput Chem 27:1787–1799
33. Bondi A (1964) J Phys Chem 64:441–451
34. Riley KE, Hobza P (2007) J Phys Chem A 111:8257–8263
35. Riley KE, Pitoňák M, Černy J, Hobza P (2010) J Chem Theor Comput 6:66–80
36. Johnson ER, Wolkow RA, DiLabio GA (2004) Chem Phys Lett 394:334–338
37. Mohan N, Vijayalakshmi KP, Koga N, Suresh CH (2010) J Comput Chem 31:2874–2882
38. Murray JS, Concha MC, Lane P, Hobza P, Politzer P (2008) J Mol Model 14:699–704
39. Politzer P, Murray JS, Bulat FA (2010) J Mol Model 16:1731–1742
40. Politzer P, Huheey JE, Murray JS, Grodzicki M (1992) J Mol Struct THEOCHEM 259:99–120
41. Politzer P, Murray JS, Lane P, Concha MC (2009) Int J Quantum Chem 109:3773–3780
42. Murray JS, Lane P, Nieder A, Klapötke TM, Politzer P (2010) Theor Chem Acc 127:345–354
43. Ignatyev IS, Schaefer HF III (2001) J Phys Chem A 105:7665–7671
44. Del Bene JE, Alkorta I, Elguero J (2010) J Phys Chem A 114:12958–12962
45. Lide DR (ed) (1997) Handbook of Chemistry and Physics, 78th edn. CRC, Boca Raton, FL
46. Brinck T, Murray JS, Politzer P (1993) Inorg Chem 32:2622–2625

Instability of C_{60} fullerene interacting with lipid bilayer

Duangkamon Baowan · Barry J. Cox · James M. Hill

Abstract Due to the large number of possible applications of nanoparticles in cosmetic and medical products, the possible hazards of nanoparticles in the human body are a major concern. A worst-case scenario is that nanoparticles might cause health issues such as skin damage or even induce cancer. As a first step to study the toxicity of nanoparticles, we investigate the energy behaviour of a C_{60} fullerene interacting with a lipid bilayer. Using the 6–12 Lennard-Jones potential function and the continuous approximation, the equilibrium spacing between the two layers of a bilayer is predicted to be 3.36 Å. On assuming that there is a circular hole in the lipid bilayer, a relation for the molecular interaction energy is determined, involving the circular radius b of the hole and the perpendicular distance Z of the spherical fullerene from the hole. A graph of the minimum energy location Z_{min} verses the hole radius b shows that a C_{60} fullerene first penetrates through a lipid bilayer when $b > 6.81$ Å, and shows a simple circular relation $Z_{min}^2 + b^2 = 6.81^2$ for Z_{min} positive and $b \leq 6.81$ Å. For $b > 6.81$, the fullerene relocates from the surface of the bilayer to the interior, and as the hole radius increases further it moves to the centre of the bilayer and remains there for increasing hole radii. Accordingly, our modelling indicates that at least for the system with no external forces, the C_{60} fullerene will not penetrate through the lipid bilayer but rather remains encased between the two layers at the mid-plane location.

Keywords Fullerenes · Interaction energy · Lennard-Jones potential · Lipid bilayer

D. Baowan
Department of Mathematics, Faculty of Science,
Mahidol University,
Rama VI,
Bangkok, Thailand

D. Baowan (✉)
Centre of Excellence in Mathematics, CHE,
Si Ayutthaya Rd.,
Bangkok 10400, Thailand
e-mail: scdbw@mahidol.ac.th

B. J. Cox · J. M. Hill
Nanomechanics Group, School of Mathematical Sciences,
The University of Adelaide,
Adelaide, South Australia 5005, Australia

B. J. Cox
e-mail: barry.cox@adelaide.edu.au

J. M. Hill
e-mail: jim.hill@adelaide.edu.au

Introduction

Nanoparticles are of considerable scientific interest and might be viewed as the connecting link between bulk materials and atomic and molecular structures. Nanoparticles may assume almost magical qualities, so that for example certain insoluble macromaterials can become soluble and certain nonconducting macrosubstances at the nanoscale can become conductors. There are many potential benefits of nanoparticles which may well outweigh any potential hazards and possible side effects [1]. Nowadays, nanoparticles are widely used in many industrial and consumer products, such as for stain and wrinkle resistant textiles or to improve cosmetics [2, 3]. However, these products raise many important health and environmental issues. For example, nanocomposites on cloth might be released during the washing process, or nanosomes in cosmetics might penetrate the skin and subsequently damage cells [2]. In this paper, we determine the interaction energy between the two layers of lipids in order to determine the optimal inter-spacing of the bilayer, and the energy profile for a C_{60} penetrating an assumed circular hole in the bilayer. Our results indicate an unusual phenomenon in that for a critical hole radius the fullerene relocates from the lipid surface to the interior for any small

increase in the hole radius. They show that in the absence of any external effects, a C_{60} fullerene is likely to remain enclosed at the mid-point between the two layers of the lipid and does not penetrate through to the other side of the lipid.

Fullerenes are an extensively studied nanomaterial because of their unique free radical chemistry and antioxidant properties [3]. Moreover, the structure of a C_{60} fullerene is simple, being a spherical surface of radius 3.55 Å and comprising 60 evenly spaced carbon atoms. Further, the ability of a fullerene to induce toxicity may require ultraviolet light and a water environment [1], and they might become dangerous to the environment and biological systems.

A lipid bilayer is very thin as compared to its lateral dimensions, and despite being only a few nanometers thick, the bilayer comprises several distinct chemical regions through its cross-section. These regions and their interactions with an aqueous environment have been characterized using x-ray reflectometry [4], neutron scattering [5] and nuclear magnetic resonance techniques [6]. The first region on either side of the bilayer is the hydrophilic head group which is typically around 8–9 Å thick. The hydrophobic core of the bilayer is typically 30–40 Å thick, but this value varies with chain length and chemistry [4, 7]. Moreover, the core thickness varies significantly with temperature, and particularly near a phase transition [8].

In terms of energy calculations, Berger et al. [9], using molecular dynamics simulations, extensively study the interaction for the bilayer of dipalmitoylphosphatidyl choline (DPPC) under various conditions, and they employ the 6–12 Lennard-Jones potential function and an electrostatic term as the force field. The translocation of C_{60} fullerenes across a DPPC lipid bilayer is studied by Qiao et al. [10] who find that the C_{60} molecules can diffuse easily into the bilayer through micropores in the membrane. Moreover, Li et al. [11] and Benrov et al. [12] use molecular dynamics to simulate the dimyristoylphoshatidylcholine (DMPC) lipid bilayer containing a fullerene, and find that the preferential position of the fullerene is about 6–7 Å away from the centre of the hydrated membrane. Further, Wong-Ekkabut et al. [13] employ a computer simulation technique to study the penetration of C_{60} fullerene clusters though a lipid membrane, and they conclude that the clusters disaggregate after entering the membrane and that high concentrations of fullerenes induce changes in the structural and elastic properties of the lipid bilayer.
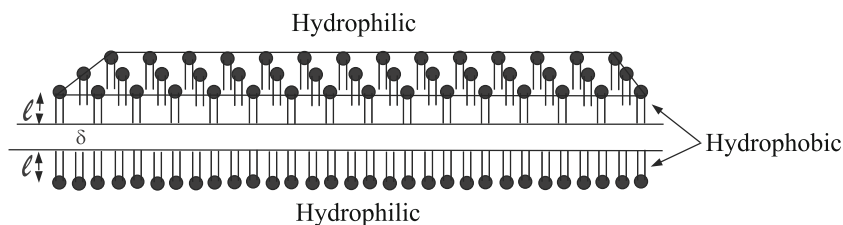
Due to the complex structure of the lipid bilayer, a number of researchers adopt a coarse grain model to study the behaviour of the lipid bilayer [13–18]. The coarse grain model smooths over or averages a number of atoms. Marrink et al. [14] describe the parametrisation of the coarse grain model for the DPPC lipid system. Further, Shelley et al. [15] use a coarse grain model to study the structure of phospholipids and conclude that this method is more efficient than Monte Carlo simulations to model the self-assembly of such systems. A new coarse-grained force field for lipids based on fitting thermodynamic and structural properties is developed by Shinoda et al. [17] who use Zwitterionic lipids as an illustration. Additionally, molecular dynamics simulations for the systems of benzene and C_{60} interacting with lipid bilayers are investigated by DeVane et al. [18] using improved cross parameters to duplicate the general behaviour expected from experiments.

In this paper, we utilise the 6–12 Lennard-Jones potential function and the continuous approximation in order to determine the interaction energy between a lipid and a C_{60} fullerene. We assume that the atoms are uniformly distributed over the entire surface of the molecules and that the molecular interaction energy can be obtained from surface or volume integrals over the molecules. A schematic of a lipid bilayer is shown in Fig. 1. We first determine the equilibrium spacing of a bilayer without a C_{60} fullerene and then we determine the interaction energy profile for a C_{60} fullerene moving through an assumed circular hole in the bilayer. In the following section, the 6–12 Lennard-Jones potential function and the continuous approximation are presented. In Inter-spacing for lipid bilayer without C_{60} fullerene, we describe the model formulation and give numerical results for the lipid bilayer without the C_{60} fullerene. On assuming a circular hole in the lipid bilayer, the energy behaviour for a C_{60} fullerene penetrating through the hole is determined and presented in Energy behaviour for C_{60} penetrating lipid bi-layer hole. Finally, we make some concluding remarks in Conclusion.

Lennard-Jones energy and continuous approximation

We employ the 6–12 Lennard-Jones potential function and the continuous approximation to calculate the molecular

Fig. 1 Schematic of lipid bilayer

Hydrophilic

Hydrophobic

Hydrophilic

interaction energy for the dual system of a lipid bilayer and a C_{60} fullerene. The Lennard-Jones function is given by

$$\Phi = -\frac{A}{\rho^6} + \frac{B}{\rho^{12}}, \tag{1}$$

where ρ denotes the distance between two typical atoms on two non-bonded molecules, and A and B are the attractive and repulsive Lennard-Jones constants, respectively. Equation 1 can also be written as

$$\Phi = 4 \in \left[-\left(\frac{\sigma}{\rho}\right)^6 + \left(\frac{\sigma}{\rho}\right)^{12} \right],$$

where $\in$ denotes the well depth and σ is the van der Waals diameter, and from which we may deduce that $A = 4 \in \sigma^6$ and $B = 4 \in \sigma^{12}$.

Using the continuous approach, the atoms at discrete locations on the molecule are assumed to be averaged over a surface or a volume, so that the molecular interaction energy for a C_{60} fullerene and a lipid bilayer is obtained by calculating integrals over the various surfaces or volumes of each molecule, and given by

$$E = \eta_1 \eta_2 \int_{S_1} \int_{S_2} \left(-\frac{A}{\rho^6} + \frac{B}{\rho^{12}} \right) dS_2 dS_1,$$

where η_1 and η_2 represent the mean surface or volume density of atoms on each molecule, and ρ here denotes the distance between two typical surface or volume elements dS_1 and dS_2. Further, we may define the integral I_n as

$$I_n = \int_{S_1} \int_{S_2} \rho^{-2n} dS_2 dS_1, \quad n = 3, 6,$$

so that, $E = \eta_1 \eta_2 (-A I_3 + B I_6)$.

The Lennard-Jones constants for the lipid bilayer are taken from the work of [14], and the head group is assumed to be a charged site (Q) interacting with an apolar carbon fullerene (C), while the tail group is assumed to be an apolar site (C) interacting with an apolar carbon fullerene (C). Further, the Lennard-Jones constants for the C_{60} fullerene in water are obtained from Bedrov et al. [12], and the numerical values are as given in Table 1. We comment that the constants A and B for the interaction between two molecular species are obtained from the accepted empirical combining laws or mixing rules [19],

given by $\in_{12} = (\in_1 \in_2)^{1/2}$ and, $\sigma_{12} = (\sigma_1 + \sigma_2)/2$, where the subscripts 1 and 2 refer to the individual chemical species.

The mean atomic surface density of the C_{60} fullerene is given by $\eta_f = 60/(4\pi a^2) = 0.3789$ Å^{-2} where a is the radius of the fullerene which is taken to be 3.55 Å. In terms of the lipid bilayer, we assume that the head group can be represented as a plane, and the tail group can be described as a rectangular box with a tail length ℓ. From the coarse grain model presented in [14], there are two interaction sites for the head group and eight interaction sites for the tail group. Further, Bedrov et al. [9] employ a value of 65 Å^2 for the area per lipid. Consequently, the mean atomic surface density for the head group is given by $\eta_{head} = 2/65 = 0.0308$ Å^{-2}, and the mean atomic volume density for the tail group can be obtained as $\eta_{tail} = 8/(65\ell) = 0.1231/\ell$ Å^{-3}, where the value of ℓ ranges from 15 to 20 Å. Table 2 presents numerical values of the various physical constants used in this paper.

In this study, we consider (I) the inter-spacing of a lipid bilayer without a C_{60} fullerene as described in Inter-spacing for lipid bilayer without C_{60} fullerene, and (II) the energy behaviour for a C_{60} fullerene moving through a circular hole on a lipid bilayer as detailed in Energy behaviour for C_{60} penetrating lipid bi-layer hole.

Inter-spacing for lipid bilayer without C_{60} fullerene

In this study, a lipid bilayer is assumed to be located on the infinite plane with a separation distance δ. Further, we assume that the head group of the lipid bilayer can be modelled as a flat plane and the tail group can be represented by a rectangular box, and the molecular interaction energy for a lipid bilayer consists of:

(i)   The interaction energy between two head groups,
(ii) Two interaction energies between head and tail   groups,
(iii) The interaction energy between two tail groups.

We begin by considering the interaction of a point with an infinite flat plane. We assume a perpendicular spacing δ between the atom and the plane, and we define a three-dimensional Cartesian coordinate system (x, y, z), in which we locate the plane of the lower surface $P(x, y) = (x, y, 0)$ and an atom $(0, 0, \delta)$, as shown in Fig. 2(a). In this case the area element of the plane is $dx\, dy$ and therefore the integral I_n becomes

$$I_n = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x^2 + y^2 + \delta^2)^{-n} dx dy. \tag{2}$$

We make the substitution $x = \sqrt{y^2 + \delta^2} \tan \psi$ which produces

$$I_n = B(n - 1/2, 1/2) \int_{-\infty}^{\infty} (y^2 + \delta^2)^{1/2 - n} dy,$$

Table 1 Numerical values of Lennard-Jones constants

Interaction	$\in$ (meV)	σ (Å)	A (eV× Å6)	B (eV× Å12)
C_{60}	4.06	3.19	17.12	1.805×10^4
Head group	18.66	4.70	804.37	8.671×10^6
Tail group	35.24	4.70	1519.37	1.638×10^7

Table 2 Numerical values of constants used in the model

Radius of C_{60} fullerene (Å)	$a=3.55$
Length of lipid tail group (Å)	$\ell = 15-20$
Mean atomic surface density for C_{60} fullerene (Å^{-2})	$\eta_f = 0.3789$
Mean atomic surface density for head group lipid bilayer (Å^{-2})	$\eta_{head} = 0.0308$
Mean atomic volume density for tail group lipid bilayer (Å^{-3})	$\eta_{tail} = 0.1231/\ell$

where $B(x, y)$ is the usual beta function. On making the further substitution $y = \delta \tan \phi$, we obtain

$$I_n = \delta^{2-2n} B(n - 1/2, 1/2) B(n - 1, 1/2) = \frac{\pi}{(n-1)\delta^{2n-2}},$$

and therefore the total interaction energy E_p between the point and the lower surface plane is given by

$$E_p = \eta_{head}\pi\left(-\frac{A}{2\delta^4} + \frac{B}{5\delta^{10}}\right), \tag{3}$$

where η_{head} is the atomic surface density of the plane P.

Now we consider the interaction energy between two flat planes, which are both assumed to be infinite in extent. The perpendicular distance between the two planes is given by $\rho = \delta$ as shown in Fig. 2(b), and the total energy can be evaluated by integrating Eq. 3 with the spacing δ, so that the integral I_n becomes

$$I_n = \frac{\pi}{(n-1)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\delta^{2n-2}} \, dx_1 dy_1,$$

which is a divergent integral. Accordingly, instead of the total interaction energy we simply calculate the interaction energy per unit area. The plane is a continuous approximation for which there is one atom occupying every $1/\eta$ of area. From these considerations, and making use of the results thus far obtained, we deduce that for two parallel planes the interaction per unit area is given by $E_{pp} = \eta_{head}E_p$, and on substitution of (3) yields

$$E_{pp} = \eta_{head}^2\pi\left(-\frac{A}{2\delta^4} + \frac{B}{5\delta^{10}}\right), \tag{4}$$

where again η_{head} is the atomic surface density of the plane which has a perpendicular separation distance δ. We note that for the purposes of determining the minimum

energy equilibrium position, we may minimise the energy per unit area E_{pp}.

On assuming that the tail group can be modelled as a box, we consider the interaction energy between the plane and a rectangular box of height ℓ which is assumed to be infinite in extent. The perpendicular distance of a point on the box to the plane is given by $\rho = z_1 + \delta$ as shown in Fig. 2(c). Therefore, we can determine the energy by integrating the expression for E_p given in (3) using this value for ρ, and the integral of interest becomes

$$I_n = \frac{\pi}{(n-1)} \int_0^\ell \frac{1}{(z_1+\delta)^{2n-2}} \, dz_1 = \frac{\pi}{(n-1)(2n-3)}\left[\frac{1}{\delta^{2n-3}} - \frac{1}{(\ell+\delta)^{2n-3}}\right].$$

On assuming that the atomic volume density of the box is η_{tail}, the interaction energy between the point and the box is given by

$$E_b = \eta_{tail}\pi\left[-\frac{A}{6}\left(\frac{1}{\delta^3} - \frac{1}{(\ell+\delta)^3}\right) + \frac{B}{45}\left(\frac{1}{\delta^9} - \frac{1}{(\ell+\delta)^9}\right)\right]. \tag{5}$$

Similarly, for the determination of the interaction per unit area for two parallel planes, the total interaction energy E_{pb} between the plane and the rectangular box per unit area is given by

$$E_{pb} = \eta_{head}\eta_{tail}\pi\left[-\frac{A}{6}\left(\frac{1}{\delta^3} - \frac{1}{(\ell+\delta)^3}\right) + \frac{B}{45}\left(\frac{1}{\delta^9} - \frac{1}{(\ell+\delta)^9}\right)\right]. \tag{6}$$

The interaction energy between two rectangular boxes can be determined either from the interaction between two planes E_{pp} or from the interaction between the plane and the box E_{pb}. Here, we integrate the expression for E_{pb} given in (6) with respect to z to obtain the energy per unit area

Fig. 2 Schematic for (**a**) atom interacting with plane P, (**b**) interaction between two flat planes, (**c**) a rectangular box of height ℓ interacting with plane P, and (**d**) interaction between two rectangular boxes, where the separation distance for all cases is denoted by δ
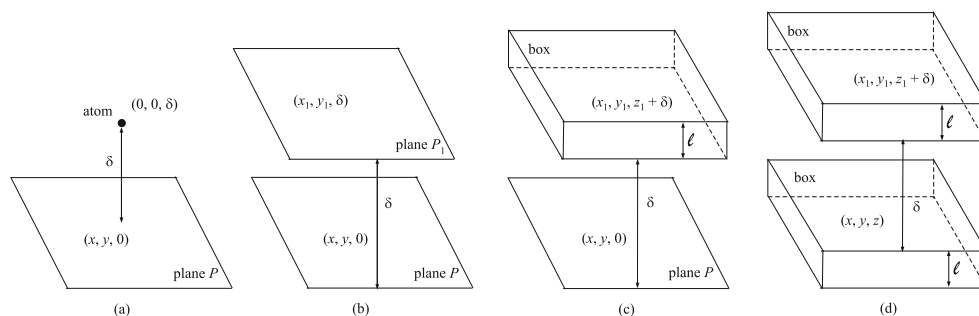
between the two boxes which are infinite in extent and of equal thickness ℓ. The perpendicular distance between a typical point in one box and the closest edge of the other box is given by $\rho = z$, and therefore, the integral I_n is of the form

$$I_n = \frac{1}{(n-1)(2n-3)} \int_\delta^{\delta+\ell} \left(\frac{1}{z^{2n-3}} - \frac{1}{(z+\ell)^{2n-3}} \right) dz.$$

This integral can be evaluated simply to yield

$$I_n = \frac{1}{2(n-1)(n-2)(2n-3)}$$
$$\times \left[\frac{1}{\delta^{2n-4}} + \frac{1}{(2\ell+\delta)^{2n-4}} - \frac{2}{(\ell+\delta)^{2n-4}} \right].$$

Thus, the interaction energy per unit area between two identical rectangular boxes with a separation distance δ can be given by
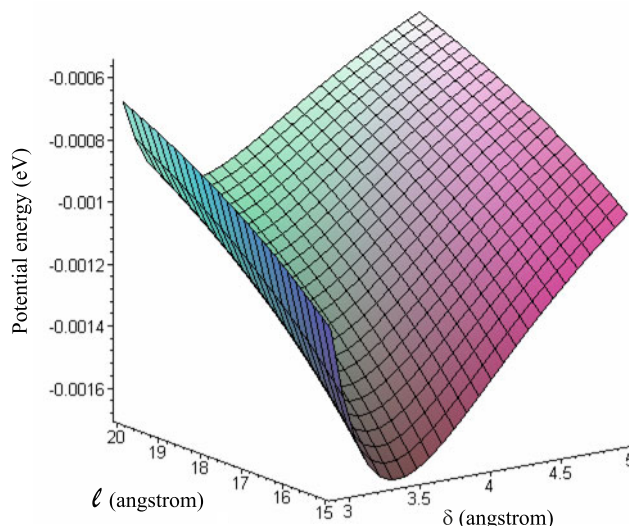
Fig. 3 Energy profile for lipid bilayer without C_{60} fullerene where δ denotes perpendicular distance between two layers and ℓ is tail length

$$E_{bb} = \eta_{tail}^2 \pi \left\{ -\frac{A}{12} \left[\frac{1}{\delta^2} + \frac{1}{(2\ell+\delta)^2} - \frac{2}{(\ell+\delta)^2} \right] + \frac{B}{360} \left[\frac{1}{\delta^8} + \frac{1}{(2\ell+\delta)^8} - \frac{2}{(\ell+\delta)^8} \right] \right\}. \tag{7}$$

We comment that the natural length for both the head and tail units is the bond length σ, while the natural energy units are those of the minimum energy ϵ. Thus for example, the head-tail interaction, Eq. 6 might be rewritten in dimensionless quantities as follows:

$$E'_{pb} = \pi N_{head} N_{tail} \left[-\frac{2}{3} \left(\frac{1}{x^3} - \frac{1}{(\lambda+x)^3} \right) + \frac{4}{45} \left(\frac{1}{x^9} - \frac{1}{(\lambda+x)^9} \right) \right],$$

where $E'_{pb} = (E_{pb}\sigma^2)/\epsilon_{pb}$ denotes an energy per unit area, $N_{head} = \eta_{head}\sigma^2$ and $N_{head} = \eta_{head}\sigma^3$ represent particle number per unit area and one per unit volume, respectively, $x = \delta/\sigma$ is a dimensionless distance between two layers, and $\lambda = 1/\sigma$ denotes a dimensionless length for the tail. This clearly indicates that the reduced length x and the constant λ are physically more relevant rather than the absolute values ℓ and δ, and the other energy expressions may be similarly interpreted.

Numerical results

Values for the molecular interaction energy of a lipid bilayer are as shown in Fig. 3 utilising the numerical values of the parameters given in Tables 1 and 2. The total potential energy of the system comprises:

(i)   One interaction energy between two head groups using E_{pp} given by (4) with a separation distance $2\ell+\delta$, and using A_{head} and B_{head},

(ii) Two interaction energies between a head group and a tail group using E_{pb} given by (6) with a separation distance $\ell + \delta$, and using the empirical mixing rule for $A_{head-tail}$ and $B_{head-tail}$,

(iii) One interaction energy between two tail groups using E_{bb} given by (7) with a separation distance δ, and using A_{tail} and B_{tail}.

We comment that the actual separation distance between the two layers of the lipid is denoted by δ as measured from the two nearest positions of the two tail layers, where ℓ is the length of the tail group. The values for A_i and B_i are those given in Table 1. Figure 3 shows the relation for the potential energy, the length of the tail ℓ, and the separation distance δ. We observe that the energy level mainly depends on the separation distance δ and it is only slightly affected by the tail length ℓ. The tail in the lipid bilayer is around 15–20 Å in length for each side. Moreover, we find that the equilibrium separation distance between the two layers of the lipid δ is 3.36 Å, which is a very small value, being ten (three) times smaller than the hydrophobic (hydrophilic) core thicknesses. In fact, the lengths of the hydrophobic tails depends on the chemical groups, and therefore the value obtained here, namely 3.36 Å, is an ideal value, and to the authors' knowledge, there are no other reported values in the literature which might be used for comparison.

Energy behaviour for C_{60} penetrating lipid bi-layer hole

Here we consider the energy behaviour for a C_{60} fullerene of radius a moving through a circular hole in a lipid bilayer as shown in Fig. 4. The lipid bilayer is assumed to be an infinite plane consisting of two head groups and two tail groups with a separation distance between the two layers of 3.36 Å as obtained from the previous section. Further, we assume that the lipid bilayer remains an equilibrium structure when interacting with the C_{60} fullerene, so that the atomic interactions between the two layers can be neglected. The atomic interaction energy between a lipid bilayer and a spherical fullerene comprises:

(i) The interaction energy for two head groups and a C_{60} fullerene,

(ii) The interaction energy for two tail groups and a C_{60} fullerene.

The head group is again modelled as a flat plane and the tail group is represented by a rectangular box. We begin by considering the molecular interaction energy for a point and a sphere, which is given by

$$E_s = \frac{\pi a \eta_f}{\rho} \left\{ \frac{A}{2} \left[\frac{1}{(a+\rho)^4} - \frac{1}{(a-\rho)^4} \right] - \frac{B}{5} \left[\frac{1}{(a+\rho)^{10}} - \frac{1}{(a-\rho)^{10}} \right] \right\},$$
(8)

and the derivation of the above equation can be found in [20]. By placing fractions over common denominators, expanding and reducing to fractions in terms of powers of $(\rho^2 - a^2)$, it can be shown that

$$\frac{A}{2\rho} \left[\frac{1}{(a+\rho)^4} - \frac{1}{(a-\rho)^4} \right] = -4aA \left[\frac{1}{(\rho^2-a^2)^3} + \frac{2a^2}{(\rho^2-a^2)^4} \right],$$
(9)

$$\frac{B}{5\rho} \left[\frac{1}{(a+\rho)^{10}} - \frac{1}{(a-\rho)^{10}} \right] = -\frac{4aB}{5} \left[\frac{5}{(\rho^2-a^2)^6} + \frac{80a^2}{(\rho^2-a^2)^7} + \frac{336a^4}{(\rho^2-a^2)^8} \right. $$
$$\left. + \frac{512a^6}{(\rho^2-a^2)^9} + \frac{256a^8}{(\rho^2-a^2)^{10}} \right].$$
(10)

The total interaction energy between the C_{60} fullerene and the head group, and that with the tail group are obtained by performing surface and volume integrals for (8) which are determined in Sects. Interaction energy between C_{60} and head group and Interaction energy between C_{60} and tail group, respectively. Further, we define the integral J_n by

$$J_n = \int_S \frac{1}{(\rho^2 - a^2)^n} dS,$$
(11)

where n is a positive integer corresponding to the power of the polynomials appearing in (9) and (10).

Interaction energy between C_{60} and head group

We define H_1 as the head group located on the xy-plane and H_2 as the other head group in the bilayer located at $z = -2\ell - \delta$. Firstly, we consider the molecular interaction energy between the head group H_1 and the C_{60} fullerene. With reference to the Cartesian coordinate system (x, y, z), the hole in the upper surface of the infinite plane is assumed to be located at $z = 0$ so that a typical point of this plane H_1 has coordinates $(x, y, 0) \equiv (r \cos \theta, r \sin \theta, 0)$ and a typical point of the hole has coordinates $(b \cos \theta, b \sin \theta, 0)$ where $r \in (b, \infty)$ and b denotes the radius of the hole. The centre of the C_{60} fullerene is assumed to be located on the axis at $(0, 0, Z)$ where Z represents the perpendicular distance from the upper surface to the centre of the fullerene, and at the mid-plane of bilayer we have $Z = -\ell - \delta/2$. Therefore, the distance from the centre of the fullerene to a typical point on the infinite plane is given by $\rho^2 = r^2 + Z^2$, and the integral J_n defined by (11) becomes

$$J_n = \int_0^{2\pi} \int_b^\infty \frac{r}{(r^2 + Z^2 - a^2)^n} dr d\theta,$$

from which we may deduce

$$J_n = \frac{\pi}{(n-1)} \frac{1}{(Z^2 + b^2 - a^2)^{n-1}}.$$

Fig. 4 Schematic for C_{60} fullerene penetrating a hole in lipid bilayer

The total interaction energy is obtained as

$$
\begin{aligned}
E_{ph} = \pi\eta_f\eta_{head}[&-4aA(J_3 + 2a^2J_4) \\
&+ \tfrac{4aB}{5}(5J_6 + 80a^2J_7 + 336a^4J_8, \\
&+ 512a^6J_9 + 256a^8J_{10})]
\end{aligned} \tag{12}
$$

and by precisely the same determination, the atomic interaction energy between the head group H_2 and the C_{60} fullerene can be obtained by substituting $Z + 2\ell + \delta_0$ for Z in J_n, where δ_0 is the equilibrium spacing between the two layers of lipid which is given by 3.36 Å.

Interaction energy between C_{60} and tail group

On assuming that the tail group can be modelled as a rectangular box, the molecular interaction energy between the two tail groups and the C_{60} fullerene can be determined. We define T_1 as the tail group connected to the head group H_1, and T_2 as the other tail group which is connected to the head group H_2. The C_{60} fullerene is centred at $(0, 0, Z)$, and a typical point of T_1 has coordinates $(r\cos\theta, r\sin\theta, -z)$ where $z \in (0, \ell)$ and ℓ is the tail length. The distance between the centre of the fullerene and the surface element of the tail group T_1 is given by $\rho^2 = r^2 + (Z + z)^2$, and the integral J_n becomes

$$
J_n = \int_0^{2\pi}\int_0^{\ell}\int_b^{\infty} \frac{r}{\left[r^2 + (Z+z)^2 - a^2\right]^n} dr\, dz\, d\theta,
$$

and we deduce

$$
J_n = \frac{\pi}{(n-1)}\int_0^{\ell} \frac{1}{\left[b^2 + (Z+z)^2 - a^2\right]^{n-1}} dz.
$$

Next we make the substitutions $z = y - Z$ and $y = \sqrt{b^2 - a^2}\tan\phi$, so that J_n becomes

$$
J_n = \frac{\pi}{(n-1)(b^2 - a^2)^{n-2}} \int_{\tan^{-1}\left(\frac{z}{\sqrt{b^2-a^2}}\right)}^{\tan^{-1}\left(\frac{z+\ell}{\sqrt{b^2-a^2}}\right)} \cos^{2n-4}\phi\, d\phi,
$$

and the above integral can be found in [21] (p.153, No.2.513.3), from which we may deduce

$$
\int \cos^{2p}\phi\, d\phi = \frac{1}{2^{2p}}\left[\binom{2p}{p}\phi + \sum_{k=0}^{p-1}\binom{2p}{k}\frac{\sin[2(p-k)\phi]}{p-k}\right], \tag{13}
$$

where $\binom{x}{y}$ is the usual binomial coefficients and $p = n - 2$. An analytical evaluation for J_n is obtained by evaluating (13) at $\phi = \tan^{-1}\left(\frac{Z}{\sqrt{b^2-a^2}}\right)$ and $\phi = \tan^{-1}\left(\frac{z+\ell}{\sqrt{b^2-a^2}}\right)$, and the total interaction energy between the tail group T_1 and the spherical fullerene is obtained as

$$
\begin{aligned}
E_{pt} = \pi\eta_f\eta_{tail}[&-4aA(J_3 + 2a^2J_4) \\
&+ \tfrac{4aB}{5}(5J_6 + 80a^2J_7 + 336a^4J_8 \\
&+ 512a^6J_9 + 256a^8J_{10})].
\end{aligned} \tag{14}
$$

The interaction energy for the tail group T_2 and the C_{60} fullerene can be obtained by precisely the same technique on replacing Z by $Z + \ell + \delta_0$ in the above integral J_n.

Numerical results

The total potential energy between the lipid bilayer with a circular hole centred at the origin and the spherical C_{60} fullerene arises from the two contributions:

(i) Two interaction energies between a head group and the C_{60} fullerene using E_{ph} given by (12), and perpendicular distances Z and $Z + 2\ell + \delta_0$ for the upper and the lower halves of the lipid bilayer, respectively, and using the empirical mixing rule for A_{f-head} and B_{f-head},

(ii) Two interaction energies between a tail group and the C_{60} fullerene using E_{pt} given by (14), and perpendicular distances Z and $Z + \ell + \delta_0$ for the upper and the lower halves of the lipid bilayer, respectively, and using the empirical mixing rule for A_{f-tail} and B_{f-tail}.

Here we employ the algebraic package MAPLE with the constant values as given in Tables 1 and 2 in order to determine the numerical results for the system. Further, the tail length ℓ is assumed to be 15 Å throughout this section. Figures 5 and 6 show the relation between the total potential energy and the perpendicular distance Z for various values of b.

In the case when the radius of the circular hole is in the range of 0–6 Å as shown in Fig. 5, the spherical fullerene is located above the lipid bilayer which corresponds to a positive value of Z, and consequently, the C_{60} fullerene does not penetrate through the bilayer. Further, we observe that the minimum energy location moves closer to the bilayer surface as the radius of the hole increases, due to the lower repulsive force from the bilayer, and that the C_{60}
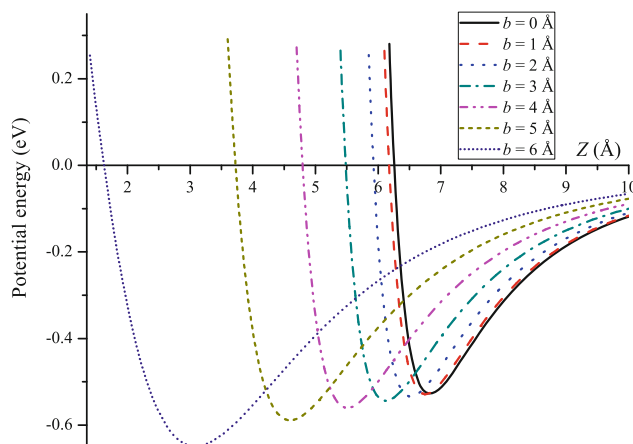
Fig. 5 Energy profile for C_{60} fullerene interacting with hole of radius $b = 0, 1, 2, 3, 4, 5, 6$ Å with respect to perpendicular distance Z with tail length ℓ assumed to be 15 Å
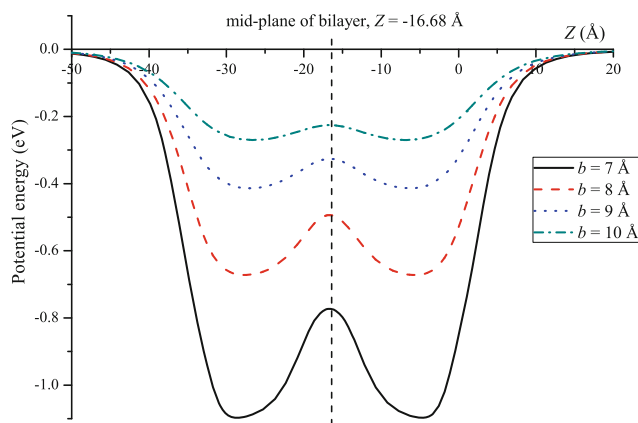
Fig. 6 Energy profile for C_{60} fullerene interacting with hole of radius $b=7, 8, 9, 10$ Å with respect to perpendicular distance Z with tail length ℓ assumed to be 15 Å

fullerene has more space to move. The centre of the C_{60} fullerene is located at the origin $Z=0$, when $b_0=6.8102$ Å. Once $b>b_0$, the fullerene penetrates through the lipid bilayer, and the energy profiles are as shown in Fig. 6. We observe two minimum energy locations due to the symmetry of the two layers of lipid. Moreover, both of the two locations indicate that the C_{60} fullerene is enclosed in the lipid bilayer. We comment that the van der Waals spacing around the C_{60} molecule is approximately 3 Å for the graphitic potential [22].

The relation between the hole radius b versus the perpendicular distance at the minimum energy Z_{min} is shown in Fig. 7. A positive value of Z_{min} indicates that the fullerene is located above the bilayer, while a negative value of Z_{min} shows that the fullerene penetrates through the surface of the bilayer. We note that the mid-plane of bilayer is given by $Z = -\ell - \delta/2$. Moreover, we observe two regions in Fig. 7 which are (i) $b \leq 6.81$ Å and (ii)
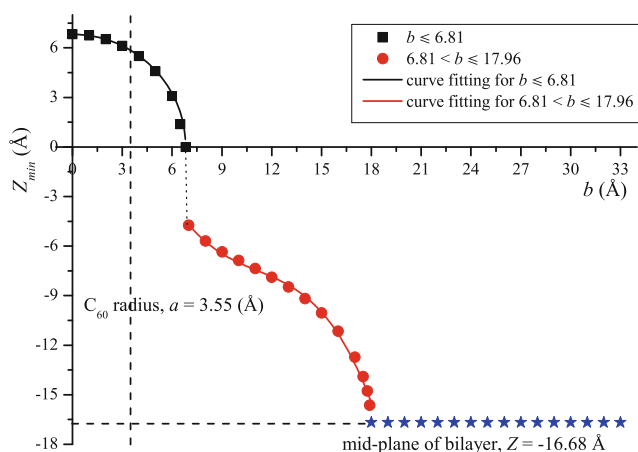
$6.81 < b \leq 17.96$ Å. On using a curve fitting technique for regions (i) and (ii), we obtain

(i): $Z_{min} = \left(6.81^2 - b^2\right)^{1/2}$,

(ii): $Z_{min} = -7.334 - 4.310 \tanh^{-1}[(b - 10.87)/7.32]$.

We observe that for region (ii), this is essentially a three parameter fit, since the 7.32 appearing in the denominator actually arises from $10.87 - a$ where $a=3.55$ Å is the C_{60} radius, and thus confirming that the curve for region (ii) asymptotes to the fullerene radius. The physical insight is that for the small b regime the problem can be thought of a hard ball of radius 6.81 Å. This length scale comes from the physical dimensions of the fullerene and the van der Waals attraction of the fullerene with the lipid bilayer. From Table 3 we see that for fullerenes of varying radii, $b - a$ has a virtually constant value of approximately 3.2 Å. Once the cutoff radius is reached the fullerene molecule relocates to a new location in the region (ii) where the molecule maximises its interaction with the upper lipid layer. As b increases further the molecule eventually finds the minimum energy located at the mid-plane of the lipid bilayer, which is $Z = -15 - (3.36/2) = -16.68$ Å.

Conclusion

In this paper, we use applied mathematical modelling to determine the molecular interaction energy and the structural dimensions of a lipid bilayer. We comment that molecular dynamics simulations of such systems are certainly possible, both atomistic on short time scales and mesoscale on longer time scales. However, such approaches while possible are computationally intensive, and the major benefit of the approach adopted here is that the calculations can be performed in a matter of seconds on standard desktop computers. Here, the 6–12 Lennard-Jones potential function for nonbonded molecules and the continuous approach are employed to determine the equilibrium spacing between two layers of the lipid. This value is found to be 3.36 Å, which to the authors' knowledge has not been reported previously in the literature.

We also assume that there is a circular hole at the centre of an infinite plane of the lipid bilayer, and the energy behaviour for a C_{60} fullerene penetrating the bilayer is

Fig. 7 Relation between minimum energy location Z_{min} and hole radius b

Table 3 Numerical values of $b - a$ for Goldberg type I fullerenes of varying radii a

	C_{60}	C_{240}	C_{540}	C_{960}	C_{1500}
a (Å)	3.5481	7.0728	10.5528	14.0342	17.5225
b (at Z=0) (Å)	6.8083	10.3116	13.7807	17.2547	20.7374
$b - a$ (Å)	3.2602	3.2388	3.2279	3.2205	3.2149

determined. We find that the minimum energy position for the C_{60} fullerene with respect to the surface of the bilayer depends on the radius b of the hole, and when $b > 6.81$ Å the fullerene penetrates through the bilayer. For a circular hole of radius $b \leq 6.81$ Å, our results indicate that a fullerene behaves like a hard sphere at rest in the hole. As the hole radius increases beyond that critical value, the fullerene relocates inside the layer until the radius acquires the value $b \leq 17.96$ Å, and for hole radii beyond that value the fullerene is attracted to the mid-plane layer and remains there. Accordingly, once the fullerene enters the lipid bilayer under no applied external force, it undergoes an instability which results in a rapid relocation to the interior of the bilayer, and then remains at the mid-point between the two layers of the lipid. In other words, for increasing hole radii, there exists an instability at the critical radius $b = 6.81$ Å.

References

1. Colvin VL (2003) The potential environmental impact of engineered nanomaterials. Nat Biotechnol 21:1166–1170
2. Thomas T, Thomas K, Sadrieh N, Savage N, Adair P, Bronaugh R (2006) Research strategies for safety evaluation of nanomaterials, part VII: Evaluating consumer exposure to nanoscale materials. Toxicol Sci 91:14–19
3. Nel A, Xia T, Mädler L, Li N (2006) Toxic potential of materials at the nanolevel. Science 311:622–627
4. Lewis BA, Engelman DM (1983) Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles. J Mol Biol 166:211217
5. Zaccai G, Blasie JK, Schoenborn BP (1975) Neutron diffraction studies on the location of water in lecithin bilayer model membranes. Proc Natl Acad Sci USA 71:376–380
6. Browning JL, Seelig J (1980) Bilayers of phosphatidylserine: a deuterium and phosphorus nuclear magnetic resonance study. Biochem 19:1262–1270
7. Rawicz W, Olbrich KC, McIntosh T, Needham D, Evans E (2000) Effect of chain length and unsaturation on elasticity of lipid bilayers. Biophys J 79:328–339
8. Trauble H, Haynes DH (1971) The volume change in lipid bilayer lamellae at the crystalline-liquid crystalline phase transition. Chem Phys Lipids 7:324–335
9. Berger O, Edholm O, Jahnig F (1997) Molecular dynamics simulations of a fluid bilayer of Dipalmitoylphosphatidylcholine at full hydration, constant pressure and constant temperature. Biophys J 72:2002–2013
10. Qiao R, Roberts AP, Mount AS, Klaine SJ, Ke PC (2007) Translocation of C60 and its derivatives across a lipid bilayer. Nano Lett 7:614–619
11. Li L, Davande H, Bedrov D, Smith GD (2007) A molecular dynamics sim ulation study of C_{60} fullerenes inside a dimyristoylphosphatidylcholine lipid bilayer. J Phys Chem B 111:4067–4072
12. Bedrov D, Smith GD, Davande H, Li L (2008) Passive transport of C_{60} fullerenes through a lipid membrand: a molecular dynamics simulation study. J Phys Chem B 112:2078–2084
13. Wong-Ekkabut J, Baoukina S, Triampo W, Tang IM, Tieleman DP, Monticelli L (2008) Computer simulation study of fullerene translocation through lipid membranes. Nat Nanotechnol 3:363–368
14. Marrink SJ, de Vries AH, Mark AE (2004) Coarse grained model for semiquantitative lipid simulations. J Phys Chem B 108:750–760
15. Shelley JC, Shelley MY, Reeder RC, Bandyopadhyay S, Moore PB, Klein ML (2001) Simulations of phospholipids using a coarse grain model. J Phys Chem B 105:9785–9792
16. Wallace EJ, Sansom MSP (2007) Carbon nanotube/detergent interactions via coarse-grained molecular dynamics. Nano Lett 7:1923–1928
17. Shinoda W, DeVane R, Klein ML (2010) Zwitterionic lipid assemblies: molecular dynamics studies of monolayers, bilayers, and vesicles using a new coarse grain force field. J Phys Chem B 114:6836–6849
18. DeVane R, Jusufi A, Shinoda W, Chiu C-C, Nielsen SO, Moore PB, Klein ML (2010) Parametrization and application of a coarse grained force field for benzene/fullerene interactions with lipids. J Phys Chem B 114:16364–16372
19. Hirschfelder JO, Curtiss CF, Bird RB (1954) Molecular theory of gases and liquids. Wiley, New York
20. Cox BJ, Thamwattana N, Hill JM (2007) Mechanics of atoms and fullerenes in single-walled carbon nanotubes. I. Acceptance and suction energies. Proc R Soc A 463:461–476
21. Gradshteyn IS, Ryzhik IM (2007) Table of integrals, series, and products 7th edn. Academic Press
22. Girifalco LA, Hodak M, Lee RS (2000) Carbon nanotubes, buckyballs, ropes, and a universal graphitic potential. Phys Rev B 62(19):13104–13110

ORIGINAL PAPER

Interactions between Al$_{12}$X (X = Al, C, N and P) nanoparticles and DNA nucleobases/base pairs: implications for nanotoxicity

Peng Jin · Yongsheng Chen · Shengbai B. Zhang · Zhongfang Chen

Abstract The interactions between neutral Al$_{12}$X(I_h) (X = Al, C, N and P) nanoparticles and DNA nucleobases, namely adenine (A), thymine (T), guanine (G) and cytosine (C), as well as the Watson−Crick base pairs (BPs) AT and GC, were investigated by means of density functional theory computations. The Al$_{12}$X clusters can tightly bind to DNA bases and BPs to form stable complexes with negative binding Gibbs free energies at room temperature, and considerable charge transfers occur between the bases/ BPs and the Al$_{12}$X clusters. These strong interactions, which are also expected for larger Al nanoparticles, may have potentially adverse impacts on the structure and stability of DNA and thus cause its dysfunction.

Keywords Aluminum cluster · DNA · Nucleobase · Base pair · Density functional theory · Interaction · Nanotoxicity

P. Jin
Wuhan Center for Magnetic Resonance,
State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences,
Wuhan 430071, China

P. Jin (✉) · Z. Chen (✉)
Department of Chemistry, University of Puerto Rico,
San Juan, PR 00931, USA
e-mail: pengjin@wipm.ac.cn

Z. Chen
e-mail: Zhongfangchen@gmail.com

Y. Chen
School of Civil and Environmental Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332, USA

S. B. Zhang
Department of Physics, Applied Physics, and Astronomy,
Rensselaer Polytechnic Institute,
Troy, NY 12180, USA

Introduction

Nanoparticles (NPs) are "fickle" and have diverse biological effects depending on their size, chemical composition, shape, surface charge density, hydrophobicity, and aggregation. These complex effects make it extremely difficult to understand nanoparticles' toxicities in different situations. As a result, nanotoxicity remains poorly understood, and a systematic approach to assessing the potential toxicology of nanomaterials is still non-existent [1, 2].

Experimentalists have made great progress (for very recent reviews, see [3–13]), especially in evaluating the toxicity of basic nanomaterials in vitro [14]. In stark contrast, computational studies aimed at understanding the toxicity of NPs are scarce [15–25], but are giving us rather instructive information. For example, by molecular dynamics (MD) simulations, Zhao et al. [16] suggested that C$_{60}$ can bind tightly to single-strand (ssDNA) or double-strand DNA (dsDNA) in aqueous solution and significantly affect the nucleotides. Zhao further found that three typical water-soluble C$_{60}$ derivatives can associate strongly with ssDNA segments and that they exhibit different binding features depending on their different functional groups [17].

Given the rapid growth of nanotechnologies and the massive production of NPs in the near future, it is impossible to test the biological effects of all the nanoscale products in vivo. Thus, fundamental theoretical research on nanotoxicity prediction is urgently required. Probing inter-

actions between NPs and biological systems is the first critical step in understanding and ultimately predicting nanotoxicity.

Aluminum, a widely used metal, caught our great attention. People are exposed to aluminum in daily life through food and water consumption as well as through the use of many commercial products, such as antacids and antiperspirants. High aluminum concentrations in the human body can lead to anemia, bone disease, and dementia [26–28]. Besides, aluminum may be associated with neurological disorders such as dialysis encephalopathy, Parkinson's dementia, and, especially, Alzheimer's disease [29, 30]. Very recent in vitro experimental studies show that Al NPs impair the cell's natural ability to respond to a respiratory pathogen regardless of NP composition (Al or Al_2O_3) [31].

Not surprisingly, theoretical and experimental studies have been performed to address the interactions between aluminum and nucleobases/base pairs or nucleotides. For example, using photoionization efficiency spectra, Pedersen et al. [32] found that Al atoms can stabilize an unusual tautomer of guanine that is incompatible with the formation of Watson-Crick base pairs (BPs). The theoretical studies of Mazzuca et al. [22] indicate that aluminum ions and nucleobases, as well as the corresponding nucleotides, can form stable complexes with large binding energies. Single aluminum atoms and trications were found to have high affinity to nucleic acid bases and monophosphate nucleotides [22, 33, 34]. However, all these studies considered only Al atoms and trications—no study involving Al NPs has been reported to the best of our knowledge.

Upon entering cells, Al NPs may coordinate with DNA and RNA nucleobases and disrupt their formation, replication and cleavage, and consequently, cause adverse effects to human health. To understand the possible toxicity of aluminum-based NPs in the human body, a detailed study on the interactions between Al NPs and DNA nucleobases as well as the corresponding BPs is urgently needed.

Choosing an appropriate NP model is crucial to the investigation of the bonding mechanism between Al NPs and DNA. Bare Al_{13} is a well-known magic cluster that has gained much attention for many years. Its neutral form is one electron deficient compared to the closed 40-electron

shell given by the jellium model [35]. The icosahedral configuration is generally considered the most energetically favorable structure of neutral Al_{13}. Substituting the inner aluminum atom of Al_{13} (I_h) with a tetravalent atom such as C (or Si) leads to a closed shell configuration, just as in a noble gas superatom [36–39]. On the other hand, doping with an atom with five valence electrons such as P (or N) results in a super alkali-metal atom with one extra electron [39, 40]. Having similar sizes and shapes but diverse redox properties, these doped $Al_{12}X$ clusters (X = Al, C, N and P) are ideal models of Al-based NPs. A comparison of these four NPs will also provide information on how the properties of NP binding to biosystems may change with different electronic configurations.

Thus, in this work, we used neutral bare $Al_{12}X$ clusters as simple NP models to investigate theoretically their interactions with adenine (A), thymine (T), guanine (G) and cytosine (C) as well as AT and GC pairs. The goal was to shed light on the possible biotoxicity caused by different NP configurations. Issues such as the geometries, electronic properties and binding Gibbs free energies of the complexes of bases/BPs with $Al_{12}X$ clusters are addressed. The mechanisms of Al NP binding as well as potential adverse effects on human biosystems are also suggested.

Computational methods

First, full geometry optimizations without symmetry constraints were carried out using the M05-2X functional [41] with the standard 6-31G* basis set. The M05-2X functional has exhibited outstanding performance in evaluating the geometries and energies of systems involving DNA bases [41, 42], and the optimized geometries of $Al_{12}X$ clusters at this level also agree well with earlier studies [43]. Several possible binding sites on nucleobases and BPs (Fig. 1, individual bases have the same numbering schemes) were considered to interact with $Al_{12}X$ (I_h) clusters, although in practice some sites may be blocked by sugar residues (ssDNA) or by intermolecular hydrogen bonds (dsDNA). All isomers were characterized as local minima by harmonic vibrational frequency analysis at the same theoretical level after the optimization. Then, the lowest-energy structures were reoptimized

Fig. 1 Geometries and numbering schemes for the Watson-Crick base pairs AT [adenine (A)–thymine (T)] and GC [guanine (G)–cytosine (C)]
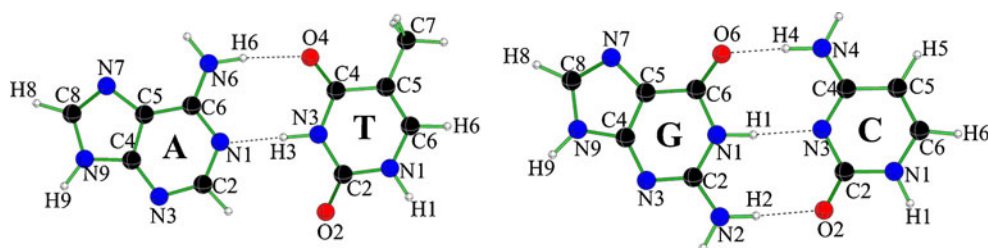
with the larger 6-311+G* basis set using polarizable continuum model (PCM) [44] to take the effect of solvent (water) into account. Charge distributions were studied with the aid of the natural bond order (NBO) analysis of Weinhold et al. [45].

To evaluate the binding strength at room temperature, we computed the binding Gibbs free energy (ΔG_b, 298.15 K and 1 atm) at the M05-2X/6-311+G* level of theory. ΔG_b is defined as the difference in Gibbs free energy between a base/BP-Al$_{12}$X complex and the separate base/BP and Al cluster (i.e., $\Delta G_b = G(\text{base/BP-Al}_{12}\text{X}) - [G(\text{base/BP}) + G(\text{Al}_{12}\text{X})]$). The computed ΔG_b values were adjusted for basis set superposition error (BSSE) using the Boys-Bernardi counterpoise correction scheme [46]. The Gaussian 03 package was employed throughout our density functional theory (DFT) computations [47]. The molecular orbitals were plotted using the gOpenmol program [48, 49].

Results and discussion

Geometries and energetics of base-Al$_{12}$X complexes

In DNA nucleobases, electron-rich N and O atoms are conventionally favored for metal binding, while the exocyclic amino groups in A, G and C can bind with metals only after deprotonation or in tautomer structures [50, 51]. Thus, we considered the initial isomers by binding Al$_{12}$X to the principal sites (either endocyclic N atoms or exocyclic carbonyl O atoms) in bases, namely, N1, N3, N7 of adenine, O2, O4 of thymine, N3, N7, O6 of guanine, and O2, N3 of cytosine. The lowest-energy complexes screened from the M05-2X/6-31G* optimized geometries (Table S1 summarizes all isomers, see Supporting Information) were reoptimized at the M05-2X/6-311+G* level of theory using PCM, and are presented in Fig. 2.

Adenine has three endocyclic nitrogen atoms (namely N1, N3 and N7, Fig. 1), all of which are potential sites for metal binding. The complexes with Al$_{12}$X bound to the N3 site of adenine are the most favorable energetically (Table S1). The nearest base-Al$_{12}$X distances (R_{b-Al}) have the order A-Al$_{13}$ (1.96 Å) < A-Al$_{12}$N (1.98 Å) < A-Al$_{12}$C (2.00 Å) = A-Al$_{12}$P (2.00 Å) (Fig. 2). The BSSE corrected ΔG_b values at 298.15 K (Table 1, −16.0, −7.3, −16.9 and −14.5 kcal mol^{-1} for A-Al$_{13}$, A-Al$_{12}$C, A-Al$_{12}$N and A-Al$_{12}$P, respectively) also suggest substantial interactions between adenine and the Al$_{12}$X. Clearly, A-Al$_{12}$C has a relatively larger ΔG_b due to the closed shell configuration of the metal cluster.

For T-Al$_{12}$X complexes, the isomers with the metal bound to the O2 atom of thymine have more favorable energies. In contrast, σ bonding to O4 was preferred in

other reports involving single Al atoms in either neutral or ionic form [22, 52]. T-Al$_{12}$X has the same ΔG_b and similar R_{b-A1} orders as A-Al$_{12}$X, i.e., for ΔG_b, T-Al$_{12}$N (−12.5 kcal mol^{-1}) < T−Al$_{13}$ (−10.9 kcal mol^{-1}) < T-Al$_{12}$P (−9.6 kcal mol^{-1}) < T-Al$_{12}$C (5.2 kcal mol^{-1}), and for R_{b-A1}, T-Al$_{13}$ (1.84 Å) = T-Al$_{12}$C (1.84 Å) < T-Al$_{12}$N (1.86 Å) < T−Al$_{12}$P (1.87 Å). However, the T-Al$_{12}$X complexes have larger ΔG_b values than A-Al$_{12}$X.

The G-Al$_{12}$X complexes binding guanine via O6 sites are all most favorable energetically. Basically, their ΔG_b values are all slightly smaller than those of A-Al$_{12}$X and T-Al$_{12}$X. G-Al$_{12}$P has the smallest ΔG_b (−19.7 kcal mol^{-1}), followed by G-Al$_{12}$N (−17.3 kcal mol^{-1}). Although relatively large, the ΔG_b of G-Al$_{12}$C is still considerable (−8.7 kcal mol^{-1}). Like that of A-Al$_{12}$X, the R_{b-Al} obeys the order of G-Al$_{13}$ (1.82 Å) < G-Al$_{12}$N (1.84 Å) < G-Al$_{12}$C (1.85 Å) = G-Al$_{12}$P (1.85 Å). Earlier studies proposed that a single Al atom or ion can bridge the O6 and the N7 atoms to form a stable structure [22, 32, 53]. As in the case of the G-Al$_{12}$X complexes studied here, the metal binding to the O6 site may distort GC pairing and lead to the formation of globular DNA [54–57].

Generally, neutral cytosine favors N3 or O2 sites for binding with metals, depending on the nature of the metal and steric factors. When an Al$_{12}$X cluster attaches to cytosine, Al$_{13}$, Al$_{12}$C and Al$_{12}$N prefer the O2 site, whereas Al$_{12}$P forms a bicoordinated complex with the N3 and O2 sites, adopting a similar binding pattern to the complex between cytosine and a single Al atom (or its cation and anion) [22, 33].

Energetically, the most stable complex is C-Al$_{12}$P (ΔG_b −21.3 kcal mol^{-1}), followed by C-Al$_{12}$N (−18.3 kcal mol^{-1}), C-Al$_{13}$ (−16.9 kcal mol^{-1}) and C-Al$_{12}$C (−9.5 kcal mol^{-1}). These complexes have the smallest ΔG_b values among all the base-Al$_{12}$X complexes studied here. A large binding energy of cytosine-Al (neutral atom) was observed experimentally and suggested theoretically [33, 34]. R_{b-Al} has the order of C-Al$_{13}$ (1.81 Å) < C-Al$_{12}$N (1.82 Å) < C-Al$_{12}$C (1.83 Å) < C-Al$_{12}$P (1.89 Å, 1.98 Å). The high stability of C-Al$_{12}$P may be due to the two-fold intra-cluster interactions (from both N–Al and O–Al bonds).

From the above, we can draw the following conclusions:

(1) Al$_{12}$X clusters prefer to bind to O atoms over N atoms in the bases. A similar behavior was also found when hard transition metals interact with DNA bases [58].

(2) Based on ΔG_b, the relative affinities of individual bases to the same Al$_{12}$X cluster are ordered as T < A < G < C. In addition, ΔG_b has the order Al$_{12}$P < Al$_{12}$N (or Al$_{12}$N < Al$_{12}$P) < Al$_{13}$ < Al$_{12}$C for the same bases. Thus, T-Al$_{12}$C has the largest ΔG_b (5.2 kcal mol^{-1}) of all the base-Al$_{12}$X complexes. The special case of C-Al$_{12}$P features bicoordination and the smallest binding Gibbs free energy (−21.3 kcal mol^{-1}).
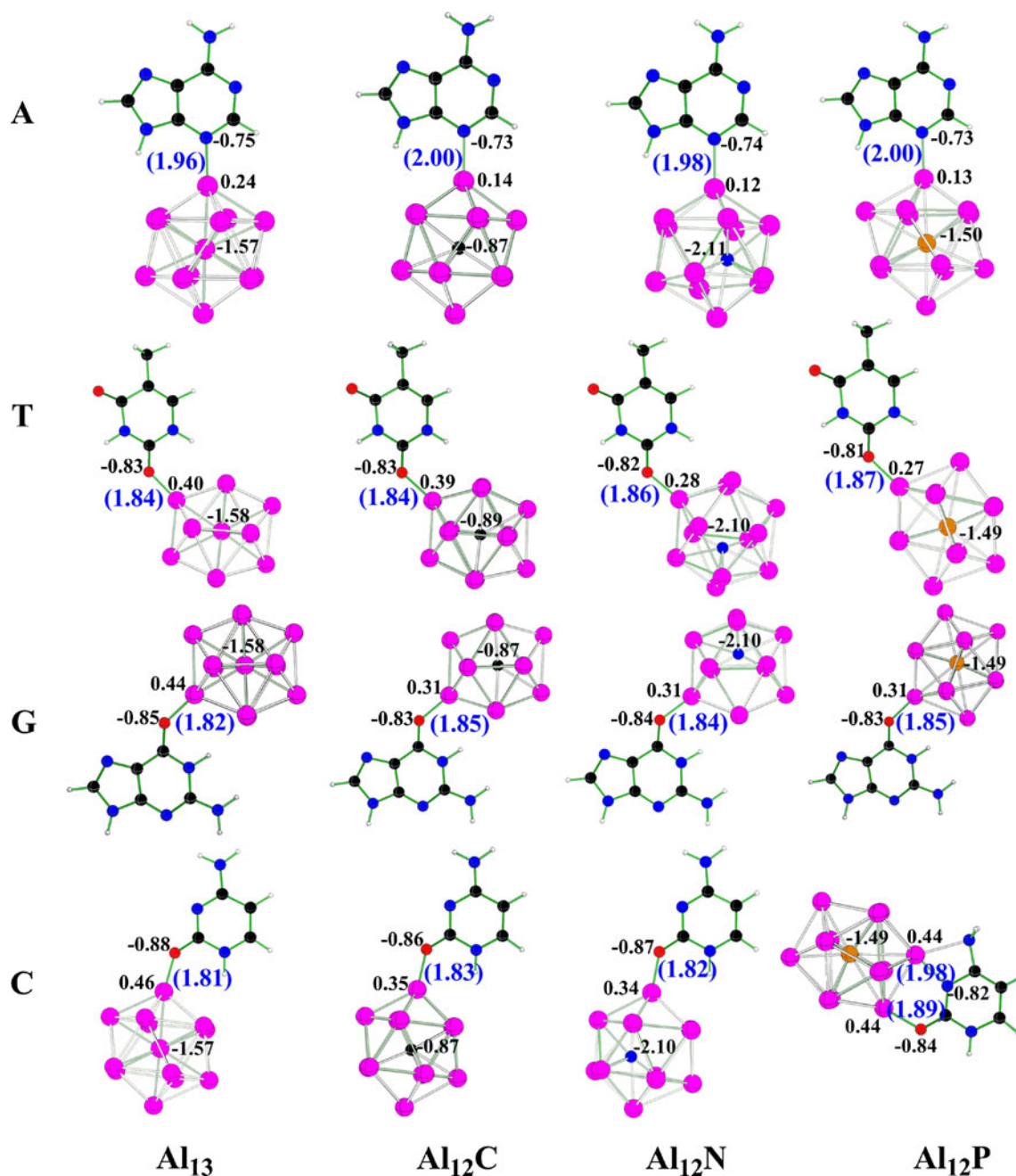
Fig. 2 Optimized geometries of the lowest-energy complexes of DNA base-Al$_{12}$X and computed natural bond order (NBO) charges for selected atoms at the M05-2X/6-311+G* level of theory. *Black* C, *red* O, *blue* N, *white* H, *pink* Al, *yellow* P. The nearest base-Al$_{12}$X distances (R_{b-Al}, unit: Å) and charges are given in *blue* (in parentheses) and *black*, respectively

(3) Basically, R_{b-Al} has the order Al$_{13}$ < Al$_{12}$N < Al$_{12}$C < Al$_{12}$P for a single base and the order C < G < T < A for a single Al$_{12}$X. Thus, C-Al$_{13}$ and A-Al$_{12}$P have the shortest (1.81 Å) and the longest R_{b-Al} values (2.00 Å), respectively, among all the base-Al$_{12}$X complexes. The shortest R_{b-Al} is attributed to the interaction between the lone pairs (from O or N atoms) on the bases and the electron-deficient outer orbitals of Al$_{13}$.

Geometries and energetics of BP-Al$_{12}$X complexes

Similar patterns are apparent in the interactions between DNA BPs and Al$_{12}$X clusters. Figure 3 illustrates the optimized structures of the complexes at the M05-2X/6-311+G* level of theory, while Table 1 summarizes the corresponding binding Gibbs free energies, ΔG_b. In the AT-Al$_{12}$X complexes, Al$_{13}$, Al$_{12}$C or Al$_{12}$N bound to the N3 site of the adenine moiety have the lowest energy, whereas Al$_{12}$P

Table 1 Basis set superposition error (BSSE)-corrected binding Gibbs free energies ΔG_b (kcal mol^{-1}) of the base/BP-Al$_{12}$X complexes computed at the M05-2X/6-311+G* level of theory

	Al$_{13}$	Al$_{12}$C	Al$_{12}$N	Al$_{12}$P
A	−16.0	−7.3	−16.9	−14.5
T	−10.9	5.2	−12.5	−9.6
G	−15.8	−8.7	−17.3	−19.7
C	−16.9	−9.5	−18.3	−21.3
AT	−14.5	−8.0	−22.3	−11.0
GC	−10.2	−4.0	−17.9	−16.2

prefers the O2 site of thymine (Table S2). In the case of GC-Al$_{12}$X complexes, however, all four Al$_{12}$X clusters prefer binding at the O6 and the N7 sites of the guanine moiety in a bridging fashion.

The BSSE-corrected ΔG_b values (−14.5, −8.0, −22.3 and −11.0 kcal mol^{-1} for AT-Al$_{13}$, AT-Al$_{12}$C, AT-Al$_{12}$N and AT-Al$_{12}$P, respectively) suggest that the interaction between AT and Al$_{12}$N is the strongest, while that between AT and Al$_{12}$C is the weakest. The R_{BP-Al} values of AT-Al$_{13}$ and AT-Al$_{12}$C are exactly the same as the R_{b-Al} values of the corresponding A-Al$_{12}$X complexes (1.96 and 2.00 Å, respectively), while the R_{BP-Al} values of AT-Al$_{12}$N (1.97 Å) and AT-Al$_{12}$P (1.82 Å) decrease by 0.01 and 0.05 Å compared to the R_{b-Al} values of A-Al$_{12}$N (1.98 Å) and T-Al$_{12}$P (1.87 Å), respectively. AT-Al$_{12}$C has the longest R_{BP-Al} and the largest ΔG_b because of the high chemical stability of Al$_{12}$C, while AT-Al$_{12}$P and AT-Al$_{12}$N

have the shortest R_{BP-Al} and the smallest binding Gibbs free energy, respectively.

With the Al$_{12}$X clusters simultaneously bridging O6 and N7 of guanine, GC-Al$_{13}$, GC-Al$_{12}$C, GC-Al$_{12}$N and GC-Al$_{12}$P all have considerable ΔG_b values (−10.2, −4.0, −17.9 and −16.2 kcal mol^{-1}, respectively). The average bond lengths (to O6 and N7 sites) are 1.94, 1.97, 1.95 and 1.93 Å, respectively, for Al$_{13}$, Al$_{12}$C, Al$_{12}$N and Al$_{12}$P, leading to the order: GC-Al$_{12}$P < GC-Al$_{13}$ < GC-Al$_{12}$N < GC-Al$_{12}$C. Among these, GC-Al$_{12}$N has the strongest binding strength and is energetically the most stable.

Intermolecular hydrogen bonds play an important role in the stabilities of BPs and their associated complexes. Thus, we computed the intermolecular hydrogen bond distances in the BP-Al$_{12}$X complexes in comparison with pure BPs (Table 2). Our computed hydrogen-bond parameters in BPs are consistent with previous high-level computations [59, 60].

For AT-Al$_{13}$, AT-Al$_{12}$C and AT-Al$_{12}$N, with respect to normal AT, the N6(A)···O4(T) hydrogen bond distances decrease slightly, by 1.3%, whereas the N1(A)···N3(T) distances increase by 2.4%. However, these two hydrogen bonds become shorter (by 3.7 and 3.5%, respectively) in the AT-Al$_{12}$P complex. Interestingly, in binding with the Al$_{12}$P, H3 (T) approaches the adenine moiety more closely [H3(T)–N1(A) and H3(T)–N3(T) are 1.05 and 1.74 Å, respectively, see Fig. 3]. In the GC-Al$_{12}$X complexes, the hydrogen bond distances of O6(G)···N4(C) increase dramatically (by 8.5, 7.4, 8.8 and 4.9% for GC-Al$_{13}$, GC-Al$_{12}$C, GC-Al$_{12}$N and GC-Al$_{12}$P, respectively), whereas most of the N1(G)···N3(C) distances are enlarged slightly
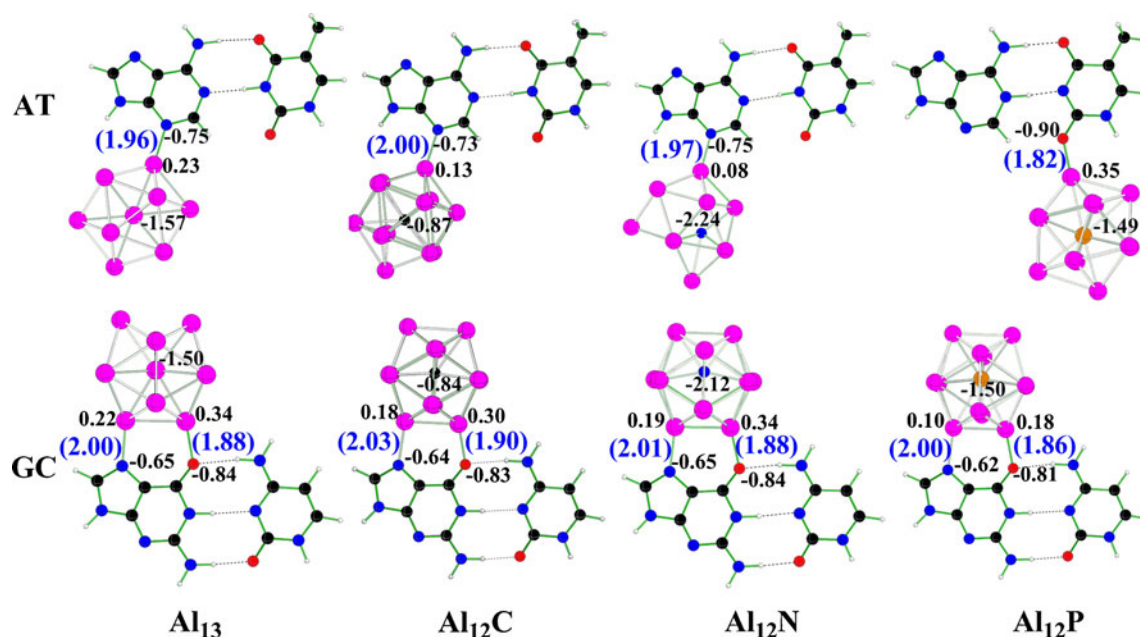
Fig. 3 Optimized geometries of the lowest-energy isomers of base pair (BP)-Al$_{12}$X complexes and computed NBO charges for selected atoms at the M05-2X/6-311+G* level of theory. Coloring and labeling scheme as in Fig. 2

Table 2 Computed hydrogen bond lengths (Å) and BSSE corrected base–base interaction energies ΔE (kcal mol^{-1}) in BP-Al$_{12}$X complexes at the M05-2X/6-311+G* level of theory[a]. *BP* Base pair

		BP	BP−Al$_{13}$	BP−Al$_{12}$C	BP−Al$_{12}$N	BP−Al$_{12}$P
AT	N6(A)···O4(T)	2.98	2.94 (−1.3)	2.94 (−1.3)	2.94 (−1.3)	2.87 (−3.7)
	N1(A)···N3(T)	2.89	2.96 (+2.4)	2.96 (+2.4)	2.96 (+2.4)	2.79 (−3.5)
	ΔE	−14.8	−14.7 (−0.7)	−14.7 (−0.7)	−14.6 (−1.4)	−97.7 (+560.1)
GC	O6(G)···N4(C)	2.84	3.08 (+8.5)	3.05 (+7.4)	3.09 (+8.8)	2.98 (+4.9)
	N1(G)···N3(C)	2.97	3.00 (+1.0)	3.01 (+1.3)	2.93 (−1.3)	2.99 (+0.7)
	N2(G)···O2(C)	2.94	2.79 (−5.1)	2.80 (−4.8)	2.83 (−3.7)	2.82 (−4.1)
	ΔE	−30.4	−31.8 (+4.6)	−31.8 (+4.6)	−30.3 (−0.3)	−31.2 (+2.6)

[a] Variations from the free BPs are given in parentheses (%)

(except GC-Al$_{12}$N). In contrast, the N2(G)···O2(C) bonds shorten considerably (by 5.1, 4.8, 3.7 and 4.1% for GC-Al$_{13}$, GC-Al$_{12}$C, GC-Al$_{12}$N and GC-Al$_{12}$P, respectively). Compared to free GC, in GC-Al$_{12}$X complexes the O6(G)···N4(C) bond is weakened and N2(G)···O2(C) is reinforced, whereas N1(G)···N3(C) changes only slightly. Clearly, depending on the nature of the metal clusters, the intermolecular H bonds of BPs can be reinforced, weakened, or even altered remarkably. This conclusion is further confirmed by the computational comparisons between the base–base interaction energies for BP-Al$_{12}$X complexes and those for the free BPs (Table 2).

Electronic properties

Charge distributions and redox properties

Detailed analyses of NBO charge distributions were performed (Table 3). Clearly, Al$_{12}$X clusters have negative charges in all complexes except C-Al$_{12}$P. Considerable charge transfers occur in all complexes, with the smallest amount of 0.16 e in T-Al$_{12}$P. Surprisingly, the largest charge transfer occurs in the C-Al$_{12}$P complex, in which the C moiety obtains a substantial amount of negative charge (−0.51 e) from the Al$_{12}$P.

We then computed the vertical and adiabatic ionization potentials (VIPs, AIPs) of the individual bases and BPs. The IP values (VIPs 8.58, 9.33, 8.30, and 9.06 eV for A, T, G and C, respectively; the corresponding AIP values are

8.32, 9.02, 7.87 and 8.92 eV) have the order G < A < C < T, which agrees very well with previous high accuracy computations [61]. Thus, in the case of AT-Al$_{12}$X, the T moiety can gain an electron from A and become negatively charged. However, in the GC-Al$_{12}$X complex, G loses an electron more easily than C, so G contributes most of the negative charge localized on the Al$_{12}$X. GC has a VIP of 7.56 eV and AIP of 7.14 eV, smaller than those of AT (VIP 8.35 eV, AIP 8.00 eV). Thus, Al$_{12}$X can obtain more electrons (ca. 0.1 e, Table 3) from GC than from AT.

In addition, we computed the vertical and adiabatic electron affinities (VEAs, AEAs) for the Al$_{12}$X clusters. The EA values (VEAs 3.23, 1.43, 2.28 and 1.44 eV, respectively, for Al$_{13}$, Al$_{12}$C, Al$_{12}$N and Al$_{12}$P; the corresponding AEA values are 3.49, 1.44, 2.28 and 1.45 eV) have the order Al$_{12}$C < Al$_{12}$P < Al$_{12}$N < Al$_{13}$. Thus, we can expect that both T and AT have small affinities to bind Al$_{12}$C, which is in line with the above discussions of ΔG_b. Moreover, we can also anticipate that Al$_{13}$ can gain some electrons from the BPs due to its large EA, which was confirmed by the computed NBO charge (Table 3), although all four Al$_{12}$X clusters obtain electrons from the BPs, regardless of whether the Al$_{12}$X cluster is electron deficient or abundant.

On the other hand, the Al atom(s) bound to the bases/BPs all exhibit positive charge (Figs. 2, 3). For example, the charges on the relevant Al atoms are 0.24, 0.14, 0.12 and 0.13 e in A-Al$_{13}$, A-Al$_{12}$C, A-Al$_{12}$N and A-Al$_{12}$P, respectively. Thus, although the whole NP can attract electrons from the base, locally the Al atom attached to

Table 3 Total NBO charges (e) in the base/BP-Al$_{12}$X complexes computed at the M05-2X/6-311+G* level of theory

	Base/BP−Al$_{13}$		Base/BP−Al$_{12}$C		Base/BP−Al$_{12}$N		Base/BP−Al$_{12}$P	
	Base/BP	Al$_{13}$	Base/BP	Al$_{12}$C	Base/BP	Al$_{12}$N	Base/BP	Al$_{12}$P
A	0.20	−0.20	0.20	−0.20	0.20	−0.20	0.19	−0.19
T	0.17	−0.17	0.17	−0.17	0.17	−0.17	0.16	−0.16
G	0.17	−0.17	0.18	−0.18	0.17	−0.17	0.17	−0.17
C	0.19	−0.19	0.19	−0.19	0.19	−0.19	−0.51	0.51
A(AT) T(AT)	0.22 −0.02	−0.20	0.20 0.00	−0.20	0.20 0.00	−0.20	0.41 −0.22	−0.19
G(GC) C(GC)	0.24 0.09	−0.33	0.24 0.08	−0.32	0.24 0.09	−0.33	0.24 0.08	−0.32

Fig. 4 Plots of the frontier orbitals for the base–Al$_{12}$X complexes and the corresponding HOMO−LUMO gap energies (eV, in parentheses) computed at the M05-2X/6-311+G* level of theory

the base is positively charged, due mainly to the connecting electronegative O or N atom.

For complexes with both bases and BPs, the charges on the inner X atoms change little for the same Al$_{12}$X (Figs. 2, 3). These complexes manifest a large electronic shielding effect of the outer Al shells. From the geometrical parameters and the charge distributions of the studied complexes, we conclude that all Al$_{12}$X NPs are coordinated
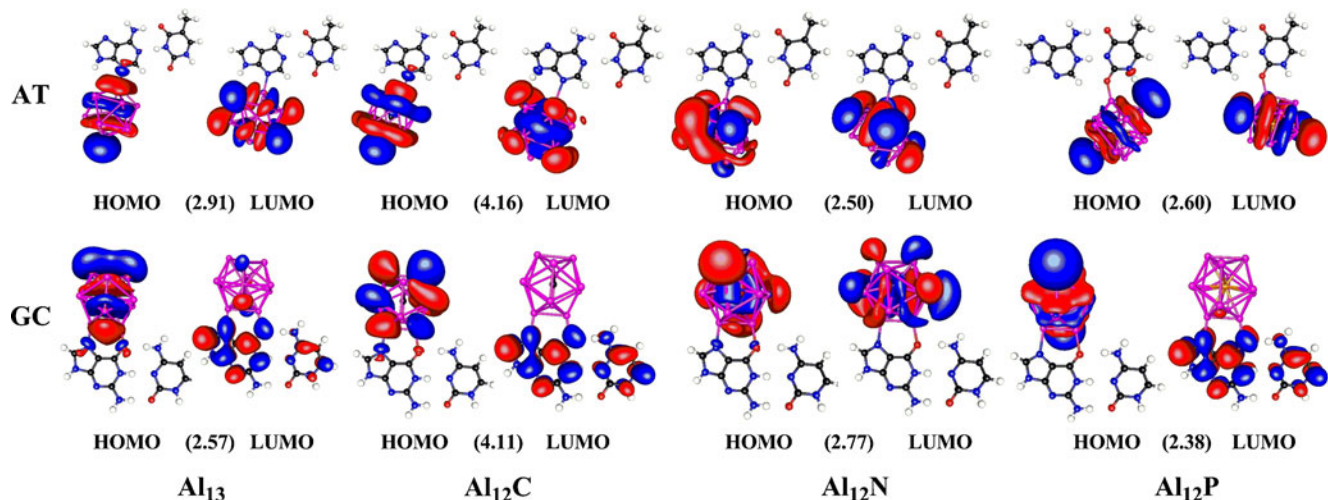
Fig. 5 Plots of the frontier orbitals for the BP-Al$_{12}$X complexes and the corresponding HOMO−LUMO gap energies (eV, in parentheses) computed at the M05-2X/6-311+G* level of theory

to bases and BPs through both covalent bonds and electrostatic interactions.

Frontier molecular orbitals

To derive useful information about the reactivity of the complexes under study, we plotted their frontier molecular orbitals, especially the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). The HOMOs and LUMOs are localized around the $Al_{12}X$ NP in almost all of the base–$Al_{12}X$ complexes (Fig. 4) with the exception of C-$Al_{12}C$, whose LUMO locates mainly on the cytosine moiety. Thus, oxidation as well as reduction reactions occur mostly on the metal clusters.

Similarly, most of the frontier orbitals of the BP-$Al_{12}X$ complexes locate on the Al clusters. The exceptions are the LUMOs in GC-$Al_{12}X$ (X = Al, C and P), which are localized mainly on the BPs (Fig. 5). Furthermore, all of the base/BP-$Al_{12}X$ complexes have sizable HOMO−LUMO gap energies (ranging from 2.58 to 4.21 eV and 2.38 to 4.16 eV for base–$Al_{12}X$ and BP-$Al_{12}X$, respectively), suggesting high kinetic stabilities.

Implications to the nanotoxicity of aluminum clusters

Our computations reveal that the $Al_{12}X$ clusters have very strong interactions with DNA bases and base pairs (To confirm this, ΔG_b values at higher temperatures (298.15–318.15 K) and BSSE-corrected binding energies were also estimated (summarized in Fig. S1 and Table S3, respectively, see Supporting Information). Notably, the typically highly stable magic cluster $Al_{12}C$ is also very reactive and interacts strongly with DNA, and the redox properties of the $Al_{12}X$ clusters do not have significant effects. It is the rather robust interaction between the surface Al atoms and the elements with high electronegativity (N and O) that binds the $Al_{12}X$ cluster and DNA. Thus, it is expected that larger Al NPs also tightly attach to DNA. Such interactions will most likely cause structural damage and electronic property changes in the DNA, and consequently lead to adverse effects on DNA function.

Conclusions

In summary, we performed a detailed theoretical study of complexes composed of DNA bases/BPs and $Al_{12}X$ (X = Al, C, N and P) clusters by means of DFT computations. All of these $Al_{12}X$ clusters attach tightly to the bases/BPs with negative binding Gibbs free energies due to the strong interactions between the outer Al atoms and the related binding sites, but the different redox properties of

$Al_{12}X$ NPs have an insignificant effect. All four NPs (one is closed shell, three are open shell) can attract electrons from the BPs. The intermolecular H bonds of the BPs can be reinforced, weakened, or altered markedly depending on the nature of the metal cluster. All these findings suggest that the bound $Al_{12}X$ clusters, as well as larger Al nanoparticles, can affect the structure and stability of DNA, and may cause negative effects on its function. We hope this study sheds light on nanotoxicology research and stimulates more theoretical studies on the toxicities of NPs.

References

1. Fischer HC, Chan WC (2007) Nanotoxicity: the growing need for in vivo study. Curr Opin Biotechnol 18:565–571
2. Lubick N (2008) Risks of nanotechnology remain uncertain. Environ Sci Technol 42:1821–1824
3. Buzea C, Pacheco II, Robbie K (2007) Nanomaterials and nanoparticles: sources and toxicity. Biointerphases 2:MR17–MR71
4. Ray PC, Yu H, Fu PP (2009) Toxicity and environmental risks of nanomaterials: challenges and future needs. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 27:1–35
5. Bouwmeester H, Dekkers S, Noordam MY, Hagens WI, Bulder AS, De Heer C, Ten Voorde SECG, Wijnhoven SWP, Marvin HJP, Sips AJAM (2009) Review of health safety aspects of nanotechnologies in food production. Regul Toxicol Pharm 53:52–62
6. Jones CF, Grainger DW (2009) In vitro assessments of nanomaterial toxicity. Adv Drug Deliv Rev 61:438–456
7. Hillegass JM, Arti S, Lathrop SA, MacPherson MB, Fukagawa NK, Mossman BT (2010) Assessing nanotoxicity in cells in vitro. Wiley Interdiscip Rev Nanomed Nanobiotechnol 2:219–231
8. Dhawan A, Sharma V (2010) Toxicity assessment of nanomaterials: methods and challenges. Anal Bioanal Chem 398:589–605
9. Hu YL, Gao JQ (2010) Potential neurotoxicity of nanoparticles. Int J Pharm 394:115–121
10. Fadeel B, Garcia-Bennett AE (2010) Better safe than sorry: Understanding the toxicological properties of inorganic nanoparticles manufactured for biomedical applications. Adv Drug Deliv Rev 62:362–374
11. Aillon KL, Xie Y, El-Gendy N, Berkland CJ, Forrest ML (2009) Effects of nanomaterial physicochemical properties on in vivo toxicity. Adv Drug Deliv Rev 61:457–466
12. Shvedova AA, Kagan VE, Fadeel B (2010) Close encounters of the small kind: adverse effects of man-made materials interfacing with the nano-cosmos of biological systems. Annu Rev Pharmacol Toxicol 50:63–88
13. Balbus JM, Maynard AD, Colvin VL, Castranova V, Daston GP, Denison RA, Dreher KL, Goering PL, Goldberg AM, Kulinowski KM, Monteiro-Riviere NA, Oberdörster G, Omenn GS, Pinkerton

KE, Ramos KS, Rest KM, Sass JB, Silbergeld EK, Wong BA (2007) Meeting report: hazard assessment for nanoparticles—report from an interdisciplinary workshop. Environ Health Perspect 115:1654–1659

14. Ostrowski AD, Martin T, Conti J, Hurt I, Harthorn BH (2009) Nanotoxicology: characterizing the scientific literature, 2000-2007. J Nanopart Res 11:251–257

15. Bosi S, Feruglio L, Ros TD, Spalluto G, Gregoretti B, Terdoslavich M, Decorti G, Passamonti S, Moro S, Prato M (2004) Hemolytic effects of water-soluble fullerene derivatives. J Med Chem 47:6711–6715

16. Zhao X, Striolo A, Cummings PT (2005) C_{60} binds to and deforms nucleotides. Biophys J 89:3856–3862

17. Zhao X (2008) Interaction of C_{60} derivatives and ssDNA from simulations. J Phys Chem C 112:8898–8906

18. Shukla MK, Leszczynski J (2009) Fullerene (C_{60}) forms stable complex with nucleic acid base guanine. Chem Phys Lett 469:207–209

19. Shukla MK, Dubey M, Zakar E, Namburu R, Czyznikowska Z, Leszczynski J (2009) Interaction of nucleic acid bases with single-walled carbon nanotube. Chem Phys Lett 480:269–272

20. Shukla MK, Dubey M, Zakar E, Namburu R, Leszczynski J (2010) Interaction of nucleic acid bases and Watson-Crick base pairs with fullerene: computational study. Chem Phys Lett 493:130–134

21. Shukla MK, Dubey M, Zakar E, Namburu R, Leszczynski J (2010) Density functional theory investigation of interaction of zigzag (7,0) single-walled carbon nanotube with Watson-Crick DNA base pairs. Chem Phys Lett 496:128–132

22. Mazzuca D, Russo N, Toscano M, Grand A (2006) On the interaction of bare and hydrated aluminum ion with nucleic acid bases (U, T, C, A, G) and monophosphate nucleotides (UMP, dTMP, dCMP, dAMP, dGMP). J Phys Chem B 110:8815–8824

23. Bedrov D, Smith GD, Davande H, Li L (2008) Passive transport of C_{60} fullerenes through a lipid membrane: a molecular dynamics simulation study. J Phys Chem B 112:2078–2084

24. Shang J, Ratnikova TA, Anttalainen S, Salonen E, Ke PC, Knap HT (2009) Experimental and simulation studies of a real-time polymerase chain reaction in the presence of a fullerene derivative. Nanotechnology 20:415101

25. Redmill PS, McCabe C (2010) Molecular dynamics study of the behavior of selected nanoscale building blocks in a gel-phase lipid bilayer. J Phys Chem B 114:9165–9172

26. Perl DP, Gajdusek DC, Garruto RM, Yanagihara RT, Gibbs CJ (1982) Intraneuronal aluminum accumuation in amyotrophic lateral sclerosis and Parkinsonism-dementia of Guam. Science 217:1053–1055

27. De Broe ME, Coburn JW (1990) Aluminium and renal failure. Dekker, New York

28. Guillard O, Fauconneau B, Olichon D, Dedieu G, Deloncle R (2004) Hyperaluminemia in a woman using an aluminum-containing antiperspirant for 4 years. Am J Med 117:956–959

29. Foncin JF (1987) Alzheimer's disease and aluminum. Nature 326:136

30. Kawahara M (2005) Effects of aluminum on the nervous system and its possible link with neurodegenerative diseases. J Alzheimers Dis 8:171–182

31. Braydich-Stolle LK, Speshock JL, Castle A, Smith M, Murdock RC, Hussain SM (2010) Nanosized aluminum altered immune function. ACS Nano 4:3661–3670

32. Pedersen DB, Simard B, Martinez A, Moussatova A (2003) Stabilization of an unusual tautomer of guanine: photoionization of Al-guanine and Al-guanine-$(NH_3)_n$. J Phys Chem A 107:6464–6469

33. Frisch MJ et al (2004) Gaussian 03, Gaussian, Inc., Wallingford CT

34. Pedersen DB, Zgierski MZ, Denommee S, Simard B (2002) Photoinduced charge-transfer dehydrogenation in a gas-phase metal-DNA base complex: Al-cytosine. J Am Chem Soc 124:6686–6692

35. De Heer WA (1993) The physics of simple metal clusters: experimental aspects and simple models. Rev Mod Phys 65:611–676

36. Khanna SN, Jena P (1992) Assembling crystals from clusters. Phys Rev Lett 69:1664–1667

37. Gong XG, Kumar V (1993) Enhanced stability of magic clusters: a case study of icosahedral $Al_{12}X$, X = B, Al, Ga, C, Si, Ge, Ti, As. Phys Rev Lett 70:2078–2081

38. Kumar V, Bhattacharjee S, Kawazoe Y (2000) Silicon-doped icosahedral, cuboctahedral, and decahedral clusters of aluminum. Phys Rev B 61:8541–8547

39. Akutsu M, Koyasu K, Atobe J, Hosoya N, Miyajima K, Mitsui M, Nakajima A (2006) Experimental and theoretical characterization of aluminum-based binary superatoms of $Al_{12}X$ and their cluster salts. J Phys Chem A 110:12073–12076

40. Wang B, Zhao J, Shi D, Chen X, Wang G (2005) Density-functional study of structural and electronic properties of Al_nN (n=2–12) clusters. Phys Rev A 72:023204

41. Zhao Y, Schultz NE, Truhlar DG (2006) Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. J Chem Theor Comput 2:364–382

42. Zhao Y, Truhlar DG (2008) Density functionals with broad applicability in chemistry. Acc Chem Res 41:157–167

43. Henry DJ, Varano A, Yarovsky I (2008) Performance of numerical basis set DFT for aluminum clusters. J Phys Chem A 112:9835–9844, and references therein

44. Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. Chem Rev 105:2999–3094

45. Reed AE, Curtiss LA, Weinhold F (1988) Intermolecular interactions from a natural bond orbital donor-acceptor viewpoint. Chem Rev 88:899–926

46. 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

47. Vázquez M-V, Martínez A (2008) Theoretical study of cytosine-Al, cytosine-Cu and cytosine-Ag (neutral, anionic and cationic). J Phys Chem A 112:1033–1039

48. Laaksonen L (1992) A graphics program for the analysis and display of molecular dynamics trajectories. J Mol Graph 10:33–34

49. Bergman DL, Laaksonen L, Laaksonen A (1997) Visualization of solvation structure in liquid mixtures. J Mol Graph Model 15:301–306

50. Lippert B (2000) Multiplicity of metal ion binding patterns to nucleobases. Coord Chem Rev 200–202:487–516

51. Hud NV (2009) Nucleic acid–metal ion interactions. RSC, Cambridge

52. Krasnokutski SA, Lei Y, Lee JS, Yang DS (2008) Pulsed-field ionization photoelectron and IR-UV resonant photoionization spectroscopy of Al-thymine. J Chem Phys 129:124309

53. Moussatova A, Vázquez M-V, Martínez A, Dolgounitcheva O, Zakrzewski VG, Ortiz JV, Pedersen DB, Simard B (2003) Theoretical study of the structure and bonding of a metal-DNA base complex: Al-guanine. J Phys Chem A 107:9415–9421

54. Robertazzi A, Platts JA (2005) Hydrogen bonding and covalent effects in binding of cisplatin to purine bases: ab initio and atoms in molecules studies. Inorg Chem 44:267–274

55. Baker ES, Manard MJ, Gidden J, Bowers MT (2005) Structural analysis of metal interactions with the dinucleotide duplex, dCG.dCG, using ion mobility mass spectrometry. J Phys Chem B 109:4808–4810

56. Pelmenschikov A, Zilberberg I, Leszczynski J, Famulari A, Sironi M, Raimondi M (1999) cis-[Pt(NH$_3$)$_2$]$^{2+}$ coordination to the N7 and O6 sites of a guanine-cytosine pair: disruption of the Watson-Crick H-bonding pattern. Chem Phys Lett 314:496–500

57. Zilberberg IL, Avdeev VI, Zhidomirov GM (1997) Effect of cisplatin binding on guanine in nucleic acid: an ab initio study. J Mol Struct THEOCHEM 418:73–81

58. Robertazzi A, Platts JA (2005) Binding of transition metal complexes to guanine and guanine–cytosine: hydrogen bonding and covalent effects. J Biol Inorg Chem 10:854–866

59. Mo Y (2006) Probing the nature of hydrogen bonds in DNA base pairs. J Mol Model 12:665–672

60. Quinn JR, Zimmerman SC, Del Bene JE, Shavitt I (2007) Does the A•T or G•C base-pair possess enhanced stability? Quantifying the effects of CH•••O interactions and secondary interactions on base-pair stability using a phenomenological analysis and ab initio calculations. J Am Chem Soc 129:934–941

61. Roca-Sanjuán D, Rubio M, Merchán M, Serrano-Andrés L (2006) Ab initio determination of the ionization potentials of DNA and RNA nucleobases. J Chem Phys 125:084302

ORIGINAL PAPER

Multiple receptor conformation docking and dock pose clustering as tool for CoMFA and CoMSIA analysis – a case study on HIV-1 protease inhibitors

Sree Kanth Sivan · Vijjulatha Manga

Abstract Multiple receptors conformation docking (MRCD) and clustering of dock poses allows seamless incorporation of receptor binding conformation of the molecules on wide range of ligands with varied structural scaffold. The accuracy of the approach was tested on a set of 120 cyclic urea molecules having HIV-1 protease inhibitory activity using 12 high resolution X-ray crystal structures and one NMR resolved conformation of HIV-1 protease extracted from protein data bank. A cross validation was performed on 25 non-cyclic urea HIV-1 protease inhibitor having varied structures. The comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) models were generated using 60 molecules in the training set by applying leave one out cross validation method, r_{loo}^2 values of 0.598 and 0.674 for CoMFA and CoMSIA respectively and non-cross validated regression coefficient r^2 values of 0.983 and 0.985 were obtained for CoMFA and CoMSIA respectively. The predictive ability of these models was determined using a test set of 60 cyclic urea molecules that gave predictive correlation (r_{pred}^2) of 0.684 and 0.64 respectively for CoMFA and CoMSIA indicating good internal predictive ability. Based on this information 25 non-cyclic urea molecules were taken as a test set to check the external predictive ability of these models. This gave remarkable out come with r_{pred}^2 of 0.61 and 0.53 for CoMFA and CoMSIA respectively. The results invariably show that this method is useful for performing 3D QSAR analysis on molecules having different structural motifs.

Introduction

Human immunodeficiency virus type1 (HIV-1) is responsible for human acquired immunodeficiency syndrome (AIDS) [1, 2], one of the most urgent world health threats. With ca. 42 million HIV/AIDS patients worldwide, only 20 anti-HIV drugs are currently available for clinical use. HIV-1 genome encodes for three major enzymes protease, reverse transcriptase and integrase for HIV-1 replication. The aspartic protease (PR) of the human immunodeficiency virus type 1 (HIV-1) cleaves the viral gag-pol fusion precursor polyprotein into active viral structural proteins and replicative enzymes such as reverse transcriptase, endonuclease, and integrase, thus playing an essential role in the maturation of HIV-1 particles and virus replication [3]. Therefore, PR is an important target for the design of specific antiviral agents dedicated to treatment of HIV-1 infection and acquired immunodeficiency syndrome (AIDS) [4–8]. cyclic urea molecules have been reported to constitute an entirely new class of potent and perspective nonpeptidic inhibitors of PR, fundamental feature of the cyclic urea inhibitors is the carbonyl oxygen that mimics the hydrogen-bonding features of the key structural water molecule present in the active site of the PR [9]. In the present study 120 cyclic urea molecules

S. K. Sivan · V. Manga (✉)
Department of Chemistry, Nizam College, Osmania University,
Hyderabad 500001, India
e-mail: vijjulathamanga@gmail.com

are taken into consideration apart from that, another 25 non-cyclic urea molecules having HIV-1 protease inhibitory activity are used for comparative molecular field analysis (CoMFA) [10] and comparative similarity indices analysis (CoMSIA) [11] based 3D QSAR analysis.

3D QSAR techniques, such as the CoMFA and CoMSIA, are based on the experimental structure-activity relationship on specific bio-macromolecule and ligand pair. This method is based only on the ligand structure and thus the spatial alignment is crucial in determining the accuracy of these approaches. Another reason for the inaccuracy in traditional ligand-based 3D QSAR [12] lies in the fact that the conformation used in the alignment may not be the active conformer of that ligand. One way to surmount such inherent imperfection is to introduce three-dimensional structures of the target bio-macromolecule during alignment process this strategy is called receptor-based 3D QSAR analysis [13–16]. The three-dimensional structures of the target can be obtained either from experimentally derived X-ray crystallography, NMR spectroscopy or from homology modeling.

Different methods have been employed for receptor-based alignment of the ligands. In general, the X-ray crystal structure of inhibitor complexed with the enzyme is used as a template for superimposition, assuming that this conformation represents the most probable bioactive conformation.

Receptor-docked alignment derived from the structure-based docking algorithms like GOLD, FlexX, GLIDE, and Autodock are used as such [17] or a rigid realignment of the poses from the receptor-docked alignment is performed, or the use of best docked mode of the smallest compound as template and modified for the other compounds [18]. These compounds were minimized and minimized structures at this binding mode were superimposed to get the molecular alignment for CoMFA and CoMSIA. To incorporate receptor flexibility, docking followed by molecular dynamics and minimization of protein ligand complex is performed and the ligand conformation obtained is used as template for alignment [19].

All these methods are helpful for the alignment of molecules having maximum common substructure, but cannot be employed for ligands having diverse structures. Here in this article we report a method to incorporate protein flexibility by applying multiple receptor conformation docking (MRCD) [20–22] and clustering of docked pose for obtaining the alignment of compounds which can have a diverse substructure.

Methodology

Twelve high resolution X-ray crystal structure and one NMR resolved crystal structure of HIV-1 protease in complex with inhibitors (pdb id: 1PRO, 1BV9, 1AJX, 1AJV, 1T7K, 1QBR, 1QBS, 1QBU, 1HVR, 1HVH, 1DMP, 1G35, 1BVG) [23–32] were downloaded from the protein data bank. GLIDE 5.6 [33] was used for molecular docking. The resolution of the crystal structure is given in Table 1. The proteins were prepared using protein preparation module applying the default parameters, grids were generated around the active site of the protease with receptor Van der Waals scaling for the non-polar atoms as 0.9 [34].

A set 145 known HIV-1 protease inhibitors that includes 120 cyclic urea and 25 non cyclic urea molecules with diverse structures and varied range of inhibition constants (Ki) were selected from literature [4, 9, 27, 28, 35–40], these were built using maestro build panel and prepared by LigPrep application in Schrödinger 2010 suite. Structures of cyclic urea molecules are given in Table 2. LigPrep produces the low energy conformer of the ligand using the MMFF94s force field. The lower energy RSSR conformations of the ligands were selected and docked into the grid generated from the 13 protein structures using the standard precision docking mode [34]. The crystal structure ligands were also docked and its RMSD was calculated to validate the docking process.

The best dock pose (low binding energy conformer according to the glide dock score) of each ligand from 13 docking runs performed on 13 receptors grids were analyzed for their hydrogen bond interactions with the receptor. The pose with the required hydrogen bonding namely with carboxylate of Asp25 and amine of Ile50 were selected for further clustering. The dock poses were clustered using clustering of conformer's script in Schrödinger 2010 suite.

Table 1 PDB ids of crystal structure of HIV-1 protease, their resolution and RMSD of redocked co-crystallized ligand

S. No.	Protein PDB id	X-ray crystal structure resolution (Å)	RMSD (Å)
1	1AJV	2.00	0.942
2	1AJX	1.85	0.497
3	1BV9	2.00	2.115
4	1DMP	2.00	0.456
5	1G35	1.80	2.05
6	1HVH	1.80	2.24
7	1HVR	1.80	0.537
8	1QBR	1.80	0.421
9	1QBS	1.80	0.521
10	1QBU	1.80	0.863
11	1T7K	2.10	0.77
12	1PRO	1.80	1.115
13	1BVG	NMR resolved	0.508

Table 2 Structures of cyclic urea molecules with their experimental pKi and predicted pKi

Mol	R^1	R^2	Expt. pKi *	Pred pKi CoMFA	Pred pKi CoMSIA
1	Metyl	Benzyl	5.240	5.307	5.328
2t	Ethyl	Benzyl	7.000	7.736	8.505
3t	n-Propyl	Benzyl	8.097	7.971	8.699
4	n-butyl	Benzyl	8.854	8.083	8.624
5	n-pentyl	Benzyl	8.796	8.951	8.902
6t	n-hexyl	Benzyl	8.337	8.359	8.433
7	2-methoxy ethyl	Benzyl	6.097	6.441	5.996
8t	2-ethoxy ethyl	Benzyl	5.958	7.130	6.493
9t	Iso-butyl	Benzyl	7.300	7.822	8.628
10t	Iso-pentyl	Benzyl	7.921	8.438	8.396
11	Iso-hexyl	Benzyl	8.155	8.346	8.219
12t	Iso-heptyl	Benzyl	7.523	8.499	8.346
13t	Allyl	Benzyl	8.284	7.726	8.633
14t	2-methylpropen-3-yl	Benzyl	8.136	7.926	8.931
15	Iso-prenyl	Benzyl	8.745	8.602	8.901
16t	(ethenyloxy)ethyl	Benzyl	7.222	7.733	8.005
17t	cyclopropyl methyl	Benzyl	8.677	8.185	8.561
18	Cyclobutyl methyl	Benzyl	8.886	8.686	8.879
19	Cyclopentyl methyl	Benzyl	8.366	8.432	8.807

The clustering was performed using atomic RMSD, in this RMSD was calculated in place which does not alter the dock pose of each conformer. The lowest binding energy conformation from the most populated cluster was chosen for CoMFA and CoMSIA analysis without further alignment, i.e., super imposition of ligands based on the common substructure for a set of molecules was not done, instead the docked conformer pose obtained form clustering were taken as is for all ligands. This imparts the flexible receptor binding information of each ligand in data set. The resulting docked pose orientations is shown in Fig. 1.

The molecules were imported into Sybyl 6.9 molecular modeling program package [41] Gasteiger-Hückel [42] charges were assigned. The standard Tripos force fields were employed for the CoMFA and CoMSIA analysis. A 3D cubic lattice of dimension 4Å in each direction with each lattice intersection of regularly spaced grid of 2.0 Å was created. The steric and electrostatic parameters were calculated in case of the CoMFA fields while hydrophobic, acceptor and donor parameters in addition to steric and electrostatic were calculated in case of the CoMSIA fields at each lattice. The sp3 carbon was used as a probe atom to generate steric (Lennard-Jones potential) field energies and

a charge of +1 to generate electrostatic (Coulombic potential) field energies. A distance dependent dielectric constant of 1.00 was used. The steric and electrostatic contributions were cut off at 30 kcal mol^{-1}.

A partial least squares (PLS) regression was used to generate a linear relationship that correlates changes in the computed fields with changes in the corresponding experimental values of biological activity (pKi) for the data set of ligands. One hundred twenty cyclic urea molecules were divided into training and test set of 60 molecules each respectively, considering the set had a balanced distribution of more and less active compounds. Twenty five non cyclic inhibitors were taken as external test set. Biological activity values of ligands were used as dependent variables in a PLS statistical analysis. The column filtering value (s) was set to 2.0 kcal mol^{-1} to improve the signal-to-noise ratio by omitting those lattice points whose energy variations were below this threshold. Cross-validations were performed by the leave-one-out (LOO) procedure to determine the optimum number of components (ONC) and the coefficient r_{loo}^2. The optimum number of components obtained is then used to derive the final QSAR model using all of the training set compounds with non-cross validation and to

Table 2 (continued)

20t	Benzyl	Benzyl	8.523	7.745	8.405
21	2-picolyl	Benzyl	6.838	6.892	6.849
22t	3-picolyl	Benzyl	8.013	7.553	8.653
23t	4-picolyl	Benzyl	7.046	8.228	8.727
24	2-napthyl methyl	Benzyl	9.508	9.471	9.467
25t	2-fluoro benzyl	Benzyl	7.468	7.695	8.521
26t	3-fluoro benzyl	Benzyl	8.523	8.195	8.892
27	4-flouro benzyl	Benzyl	8.854	8.797	8.666
28	3-chlorobenzyl	Benzyl	9.051	8.667	8.445
29t	4-chloro benzyl	Benzyl	8.284	7.840	8.121
30t	3-bromo benzyl	Benzyl	8.854	8.412	8.158
31	4-bromobenzyl	Benzyl	7.569	7.762	7.745
32t	3-methyl benzyl	Benzyl	8.155	8.668	8.387
32	4-methyl benzyl	Benzyl	8.244	8.367	8.427
34	3-(trifluoro methyl) benzyl	Benzyl	7.657	7.732	7.580
35	4-(trifluoro methyl) benzyl	Benzyl	7.292	7.135	7.383
36t	3-methoxy benzyl	Benzyl	8.796	9.141	9.041
37	3-(hydroxy methyl) benzyl	Benzyl	9.854	9.582	9.776
38t	4-hydroxy benzyl	Benzyl	9.921	8.485	9.267
39	3-carbamoyl benzyl	Benzyl	10.409	10.637	10.551
40	3-(hydrazinyl carbonyl) benzyl	Benzyl	10.745	10.754	10.791
41t	2-(ethyl carbamoyl) benzyl	Benzyl	9.678	9.216	9.718
42t	2-(isopropyl carbamoyl)	Benzyl	9.237	8.720	9.273
43	3-(propylcarbamoyl) benzyl	Benzyl	9.445	9.557	9.340
44	3-(butylcarbamoyl) benzyl	Benzyl	9.373	9.233	9.284
45	3-[(cyclopropylmethyl) carbamoyl] benzyl	Benzyl	9.130	9.039	9.006
46t	2-[(2,2,2-trifluoroethyl) carbamoyl]benzyl	Benzyl	9.678	9.930	10.085
47	3-(cyanocarbamoyl) benzyl	Benzyl	10.201	10.390	10.266
48	3-(phenylcarbamoyl) benzyl	Benzyl	9.366	9.390	9.379
49	3-(pyridin-4-yl carbamoyl)benzyl	Benzyl	9.387	9.348	9.316
50	3-(pyridin-3-ylcarbamoyl)benzyl	Benzyl	9.538	9.529	9.624
51t	2-(pyridin-2-yl carbamoyl)benzyl	Benzyl	10.367	10.048	9.262
52	3-[(5-methylpyridin-2-yl)carbamoyl]benzyl	Benzyl	10.959	11.019	10.944
53t	2-[(6-methylpyridin-2-yl)carbamoyl]benzyl	Benzyl	10.699	9.586	10.283
54t	2-[(5-chloropyridin-2-yl)carbamoyl]benzyl	Benzyl	10.921	9.785	9.262
55t	2-[(3,5-dichloropyridin-2-yl)carbamoyl]benzyl	Benzyl	9.610	9.658	9.107
56	3-[(5-bromopyridin-2-yl)carbamoyl]benzyl	Benzyl	10.456	10.665	10.465
57	3-[(4-methylpyrimidin-2-yl)carbamoyl]benzyl	Benzyl	9.939	9.805	9.979
58	3-(pyrimidin-2-ylcarbamoyl) benzyl	Benzyl	10.745	10.700	10.890
59t	2-(pyrimidin-2-ylcarbamoyl)benzyl	Benzyl	9.818	9.891	11.018
60t	2-(1,3-thiazol-2-ylcarbamoyl)benzyl	Benzyl	10.569	10.467	10.237

obtain the conventional regression coefficient (r^2). Since the statistical parameters were found to be the best for the model from the LOO method, it was employed for further predictions of activity of test molecules for cross validation of the model. The schematic representation of the multiple receptor conformation docking (MRCD), clustering and 3D QSAR is given in Fig. 2.

Multiple receptors conformation docking is a method employed for incorporating the receptor flexibility while docking analysis. In this method different conformations of

Table 2 (continued)

61t	2-[(5-methyl-1,3-thiazol-2-yl)carbamoyl] Benzyl	Benzyl	10.602	9.115	9.070
62t	2-[(4-methyl-1,3-thiazol-2-yl) carbamoyl]benzyl	Benzyl	10.854	10.426	9.894
63	3-(1*H*-benzimidazol-2-ylcarbamoyl)benzyl	Benzyl	10.620	10.537	10.643
64	3-{[5-(trifluoromethyl)-1,3,4-thiadiazol-2-yl] carbamoyl}benzyl	Benzyl	9.745	9.874	9.684
65	3-(1,3,4-thiadiazol-2-ylcarbamoyl) benzyl	Benzyl	9.959	9.984	9.962
66t	3-(1*H*-pyrazol-5-yl) benzyl	Benzyl	10.569	9.891	11.424
67	[3-(ethylamino)-1*H*-indazol-6-yl]methyl	Benzyl	10.387	10.357	10.268
68t	[3-(propan-2-ylamino)-1*H*-indazol-5-yl]methyl	Benzyl	10.046	10.637	10.909
69t	{3-[(cyclopropyl methyl) amino]-1*H*-indazol-5-yl}methyl	Benzyl	9.469	9.925	8.756
70	[3-(propylamino)-1*H*-indazol-6-yl]methyl	Benzyl	9.638	9.648	9.479
71	3-(2-hydroxypropan-2-yl) benzyl	Benzyl	10.100	10.318	10.094
72t	1*H*-indazol-5-yl ethyl	Benzyl	10.745	9.517	9.786
73t	(3-methoxy-1*H*-indazol-5-yl)methyl	Benzyl	9.750	9.636	9.870
74	3-[(aminoacetyl) amino] benzyl	Benzyl	10.530	10.451	10.513
75	Benzyl	Methyl	5.301	5.223	5.305
76t	Benzyl	4-isopropyl benzyl	8.959	8.840	9.315
77	Benzyl	4-(methylthio)benzyl	8.469	8.346	8.304
78	Benzyl	Iso-butyl	5.770	6.067	6.412
79t	Benzyl	2-(methyl thio) ethyl	5.959	6.665	6.962
80t	Benzyl	Cyclohexyl methyl	7.553	8.257	7.986
81t	Benzyl	Phenethyl	6.495	7.162	7.917
82	Benzyl	(napthalene-2-yl) methyl	8.009	7.977	8.144
83	Benzyl	(4-methyl sulfanyl) benzyl	8.602	8.519	8.455
84	Benzyl	(4-methyl sulfonyl) benzyl	8.602	8.577	8.540
85t	Benzyl	2-methoxy benzyl	7.222	8.914	8.275
86t	Benzyl	2-hydroxy benzyl	7.456	8.164	8.261
87	Benzyl	3-methoxy benzyl	8.328	8.493	8.300
88t	Benzyl	4-methoxy benzyl	8.066	8.606	8.552
89t	Benzyl	4-hydroxy benzyl	8.959	8.251	8.139
90	Benzyl	3-amino benzyl	8.553	8.596	8.447
91	Benzyl	(3-dimethyl amino) benzyl	8.367	8.514	8.514
92	Benzyl	4-amino benzyl	8.071	8.220	8.109
93t	Benzyl	4-picolyl	7.658	7.214	7.843
94	Benzyl	(1,3-benzodioxol-5-yl) methyl	8.886	8.691	8.697
95t	Benzyl	4-(2-hydroxyethoxy)benzyl	9.119	8.924	9.447
96t	Benzyl	4-(pyridin-4-ylmethoxy)benzyl	9.076	8.877	8.924
97	Cyclopropyl methyl	Iso-butyl	7.066	7.171	7.113
98	Cyclopropyl 1 methyl	Iso-propyl	6.602	6.613	6.263
99t	Cyclopropyl methyl	(3-methyl sulfanyl) propyl	5.602	7.676	7.100

Table 2 (continued)

100t	Cyclopropyl methyl	4-fluorobenzyl	8.237	8.047	8.300
101t	Cyclopropyl methyl	3-methoxy benzyl	9.060	8.662	8.332
102	Cyclopropyl methyl	3-hydroxy benzyl	7.886	8.045	7.751
103	Cyclopropyl methyl	4-methoxy benzyl	8.538	8.756	8.804
104t	Cyclopropyl methyl	2-naphthylmethyl	8.367	8.675	9.190
105t	Cyclopropyl methyl	2-thienyl methyl	8.041	7.392	7.483
106t	4-hydroxy benzyl	2-(methylthio) ethyl	5.409	7.101	7.007
107t	4-hydroxy benzyl	Cyclohexyl methyl	7.495	8.443	8.757
108t	4-hydroxy benzyl	4-fluorobenzyl	9.357	8.468	8.895
109t	4-hydroxy benzyl	3-methoxy benzyl	9.959	9.305	9.364
110t	4-hydroxy benzyl	3,4 difluoro benzyl	9.328	8.205	8.885
111	4-hydroxy benzyl	Pyridine-4-yl methyl	8.319	8.126	8.407
112	4-hydroxy benzyl	4-methoxy benzyl	9.620	9.672	9.512
113	4-hydroxy benzyl	Iso-butyl	7.420	7.119	7.299
114t	3-butanoyl benzyl	Phenyl	8.860	8.204	8.354
115	3-(2,2-dimethyl propanoyl)benzyl	Phenyl	8.450	8.391	8.524
116	3-(1*H*-imidazol-5-ylmethyl)benzyl	Phenyl	9.730	9.831	9.694
117t	3-(1*H*-benzimidazol-2-yl carbamoyl)benzyl	Ethyl	8.347	8.462	7.958
118	3-(1*H*-benzimidazol-2-yl carbamoyl)benzyl	4-amino benzyl	10.796	10.660	10.836
119	3-(1*H*-pyrazol-3-yl) benzyl	4-hydroxy benzyl	10.796	10.876	10.854
120	3-(1*H*-pyrazol-3-yl) benzyl	4-methoxy benzyl	10.108	9.993	10.211

the same protein are taken and ligands are docked into the grids generated from these conformations. The accuracy of a docking procedure lies in how closely the lowest energy pose (binding conformation) predicted by the object scoring function (Glide score), resembles an experimental binding mode as determined by X-ray crystallography. In the present study, standard precision glide docking procedure was validated by removing crystal structure ligand from the binding site and redocking it to the binding site of HIV-1 protease for each of the receptor conformations.

Results and discussion

We found a very good agreement between the localization of the inhibitor upon docking and from the crystal structure, i.e., having similar hydrogen bonding interactions with Asp 25 and Ile 50. The root mean square deviations between the predicted conformation and the observed X-ray crystallographic conformation for the ligands ranged from 0.4Å to 2.24Å, the values are provided in Table 1, these values suggests the reliability of Glide docking in
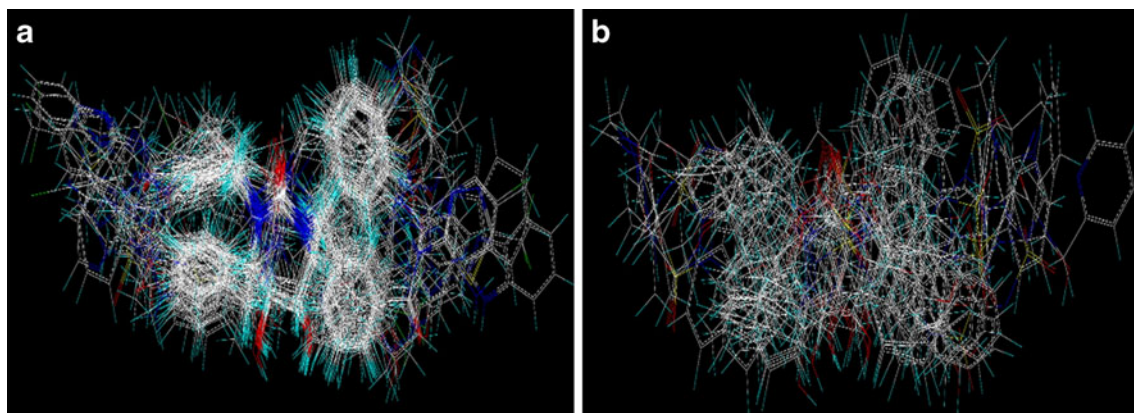
Fig. 1 (**a**) clustered pose of cyclic urea molecules from MRCD (**b**) clustered pose of non-cyclic urea molecules from MRCD
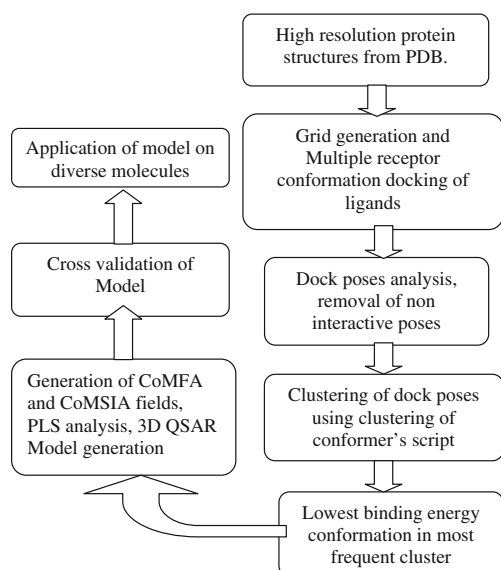
Fig. 2 Schematic representation of the multiple receptor conformation docking, clustering and 3D QSAR

reproducing the experimentally observed binding mode for HIV-1 protease inhibitor and the parameter set for the Glide docking is reasonable to reproduce the X-ray structure.

The advantage of MRCD is clearly understood when each molecules dock pose was analyzed to confirm the presence of required hydrogen bond interaction with the active site amino acids namely with carboxylate of Asp25 and amine of Ile50. In some receptor confirmation the interactions were missing or inverse, in this manner the redundant poses of the molecules can be screened out and the biologically active conformation of the ligand can be obtained. To obtain the most preferred biologically active conformer of the ligand, a cluster analysis of the dock poses of each ligand was performed using clustering of conformer's script. Pose from the most frequent cluster having lowest binding energy was selected. This pose of each molecule were taken as the basis for the CoMFA and CoMSIA analysis, these conformation provide the most vital information for the binding of ligand into the protein active site.

3D QSAR analysis was done by dividing the molecules into training and test set having 60 molecules each, keeping in view that the activity range is at least 5 log units different in both the sets. The CoMFA and CoMSIA statistical analysis is summarized in Table 3. Statistical data shows r_{loo}^2 0.598 for CoMFA and 0.674 for the CoMSIA models, respectively, which indicates a good internal predictive ability of both models. The models developed also exhibited r^2 of 0.983 and 0.985 for CoMFA and CoMSIA, respectively. To test the predictive ability of the models, a test set of 60 molecules excluded from the model derivation was used. The predictive correlation coefficient r_{pred}^2 of

Table 3 Summary of CoMFA and CoMSIA statistical analysis

Statistical parameters	CoMFA	CoMSIA
PLS result summary for model derived from cyclic urea molecules		
r_{loo}^2 [a]	0.598	0.674
Number of cyclic urea molecules in training set	60	60
Number of cyclic urea molecules in test set	60	60
Number of non-cyclic urea molecules in test set for external validation	25	25
ONC [b]	8	10
SEE [c]	0.199	0.189
r^2 [d]	0.983	0.985
F_{ratio} [e]	153.166	311.772
r_{pred}^2 [f]	0.684	0.640
r_{pred}^2 on non-cyclic urea molecules	0.61	0.53
Fraction of field contributions		
Steric	56.2	14.2
Electrostatic	43.8	28.8
Hydrophobic	–	23.3
Acceptor	–	18.6
Donor	–	14.8
PLS result summary for model derived from non cyclic urea molecules		
r_{loo}^2 [a]	0.595	0.568
Number of non-cyclic urea molecules in training set	25	25
Number of cyclic urea molecules in test set	60	60
ONC [b]	10	10
SEE [c]	0.048	0.076
r^2 [d]	0.997	0.998
F_{ratio} [e]	225.814	895.159
r_{pred}^2 [f]	0.42	0.41
Fraction of field contributions		
Steric	57.8	14.9
Electrostatic	42.2	21.7
Hydrophobic	–	17.7
Acceptor	–	20.2
Donor	–	25.5

[a] correlation coefficient from leave one out method

[b] optimum number of components

[c] standard error of estimate

[d] conventional regression coefficient

[e] Fisher test value

[f] predictive r^2 on test set using equation ($r^2 =$(SD-PRESS)/SD) where SD is the sum of the squared deviations between the biological activities of the test molecules and the mean of training set. PRESS is the sum of the squared deviation between the observed and the predicted activities of the test set

0.684 for CoMFA and 0.640 for CoMSIA models indicate good external predictive ability of the model. The experimental and predicted activity from CoMFA and CoMSIA model is given in Table 2.

Based on this information 25 non-cyclic urea molecules were taken as a test set to check the external predictive ability of these models on diverse set of molecules (structure are given in Fig. 3). This gave remarkable outcome with r_{pred}^2 of 0.61 and 0.53 for CoMFA and CoMSIA respectively. The results invariably show that this method is useful for performing 3D QSAR analysis on molecules having different structural motifs. The experimental and predicted activity is given in Table 4. Further, these diverse structures were used to generate a QSAR model, that gave statistical data of r_{loo}^2 0.595 for CoMFA and 0.568 for the CoMSIA models respectively, gave an r^2 of 0.997 and 0.998 for CoMFA and CoMSIA, respectively. The graph for the experimental and predicted pKi values for training set, test set and external data set of non-cyclic urea molecules are shown in Fig. 4. The 60 cyclic urea molecule of the training set was used to check the predictive ability of this model; it showed an acceptable but not very enthusiastic result of 0.42 and 0.41 for CoMFA and CoMSIA respectively. The probable reason for this result could be the smaller training set of 25 non cyclic urea

molecules and larger test set of 60 cyclic urea molecules. By using the QSAR ANALYSIS LIST command the intercept for the PLS equations were obtained, the command does not provide the regression equation, hence a method followed by Wheelock and Nakagawa et al. [43] was used to obtain the QSAR equation where the intercept was used. The equations derived for the CoMFA and CoMSIA analysis is provided below (Eqs. 1, 2, 3, 4) where n is number of molecules involved in the model generation and r is the correlation coefficient for the equations obtained.

PLS derived from cyclic urea molecules

$$pKi = [CoMFA\ terms] + 5.118\ n = 60,\ r = 0.991 \quad (1)$$

$$pKi = [CoMSIA\ terms] + 6.214\ n = 60,\ r = 0.992 \quad (2)$$

PLS derived from non-cyclic urea molecules

$$pKi = [CoMFA\ terms] + 5.193\ n = 25,\ r = 0.965 \quad (3)$$

$$pKi = [CoMSIA\ terms] + 6.192\ n = 25,\ r = 0.936 \quad (4)$$

The contour generated from the above two QSAR models for cyclic and non-cyclic urea molecules using docking and clustering as a prerequisite gave results that invariably reveal that this method is useful for performing 3D QSAR analysis on molecules having different structural motifs.
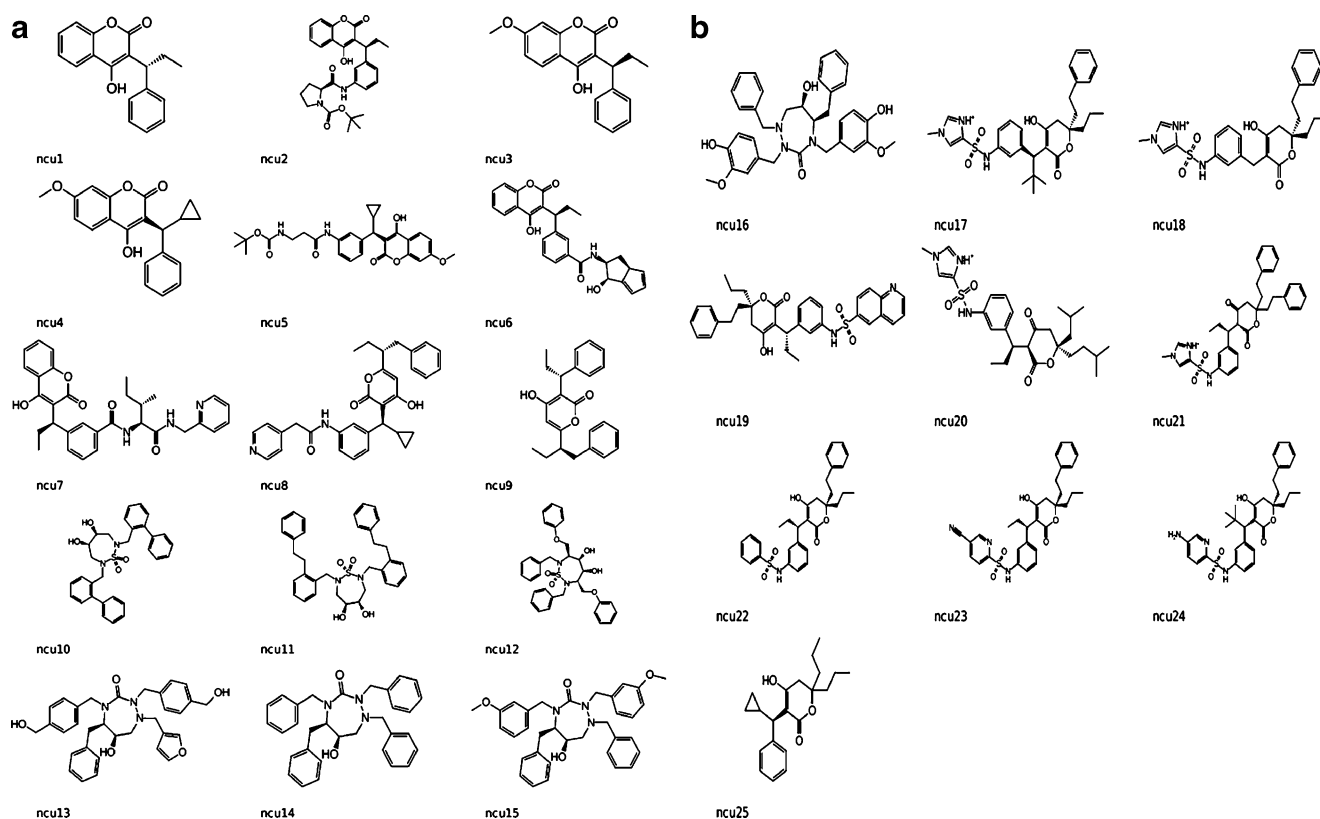
Fig. 3 Structure of non-cyclic urea molecule

Table 4 Experimental and predicted pKi values of non-cyclic urea molecules

Mol.	Experimental pKi	Predicted pKi CoMFA cyclic urea PLS	Predicted pKi CoMSIA cyclic urea PLS	Predicted pKi CoMFA non cyclic urea PLS	Predicted pKi CoMSIA non cyclic urea PLS
NCU 1	6	5.648	6.972	5.993	5.958
NCU 2	5.44	7.685	8.186	5.425	5.462
NCU 3	6.24	6.353	7.199	6.323	6.459
NCU 4	6.77	6.267	7.194	6.683	6.583
NCU 5	7.55	8.972	8.501	7.505	7.542
NCU 6	5.47	7.145	8.282	5.486	5.465
NCU 7	6.4	7.484	9.042	6.405	6.375
NCU 8	7.36	7.021	7.216	7.381	7.346
NCU 9	7.42	7.915	7.699	7.748	7.386
NCU 10	5.6	7.797	8.116	5.614	5.599
NCU 11	8.52	8.217	8.015	8.512	8.506
NCU 12	7.72	8.691	8.183	7.722	7.73
NCU 13	9.31	7.614	9.062	9.305	9.3
NCU 14	8.7	7.995	8.306	8.66	8.716
NCU 15	9.66	9.081	9.216	9.713	9.629
NCU 16	11.3	9.129	10.062	11.3	11.331
NCU 17	8.92	8.616	8.908	8.002	7.085
NCU 18	7.89	7.9	8.668	7.941	7.803
NCU 19	8.51	8.531	8.861	8.141	8.315
NCU 20	8.52	9.042	9.716	8.528	8.523
NCU 21	9	10.257	8.634	9.002	8.994
NCU 22	7.77	7.62	7.393	8.148	7.617
NCU 23	8	7.373	8.005	8.07	8.241
NCU 24	9.7	8.418	9.447	8.315	7.861
NCU 25	7.82	7.829	8.336	7.01	7.383

The contour maps of CoMFA (electrostatic and steric) and CoMSIA (electrostatic, steric, hydrophobic, donor and acceptor) are represented by color codes. The contour maps of CoMFA denote the region in the space where the aligned molecules would favorably or unfavorably interact with the receptor while the CoMSIA contour maps denote those areas within the specified region where the presence of a group with a particular physicochemical activity binds to the receptor. The CoMFA/CoMSIA results were graphically interpreted by field contribution maps using the 'STDEV * COEFF' field type.

Steric and electrostatic contour maps

To visualize the information content of the derived 3D QSAR models, CoMFA contours maps were generated to rationalize the regions of 3D space around the cyclic urea and non-cyclic urea molecules, where changes in the steric and electrostatic fields would influence the increase or decrease in inhibitory activity. All of the contours represented the default 80 and 20% level contributions for favored and disfavored regions, respectively. The

CoMFA steric and electrostatic contour maps are shown in Figs. 5 and 6 respectively.

The steric field is characterized by green and yellow contours, in which green contours indicate the region where bulkier group would be favorable, while the yellow represents region were bulkier group would decrease the activity. The most potent cyclic urea analogue from the series, compound 52 was embedded in the map (Fig. 5a), and compound NCU16 (Fig. 5b) was embedded for non-cyclic urea molecule, to demonstrate their affinity for the steric regions of inhibitors. The steric contours for both cyclic and non-cyclic urea molecules showed same regions, having a large green contour at the R^2 and R^1 positions of the cyclic urea molecule suggesting an increase in the bulkiness would increase the activity of the molecules, similar kind of contour was obtained for non-cyclic urea molecule where the green contour was envisaged at the same position as in cyclic urea. Substitution on the phenyl ring at R^1 would decrease the activity in cyclic urea molecules, a similar kind of contour was observed for non-cyclic urea where the yellow contours is observed all over the molecule expect at R^2 position
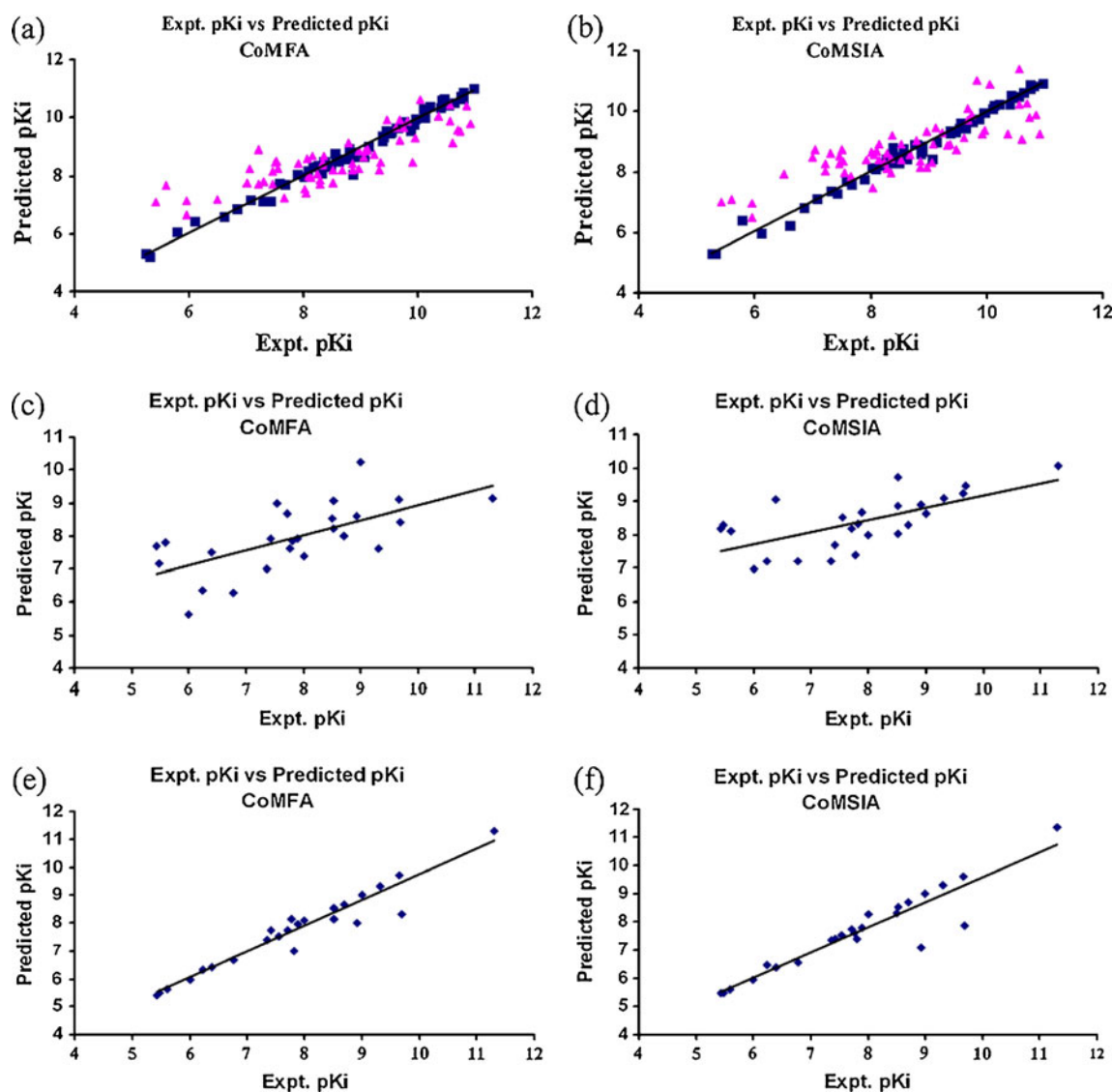
Fig. 4 (**a**) and (**b**) Scatter plot of experimental vs predicted pKi values for cyclic urea molecules using (PLS) derived from cyclic urea set (test set is represented in triangles), (**c**) and (**d**) Scatter plot of experimental vs predicted pKi values for non-cyclic urea molecules using (PLS) derived from cyclic urea set, (**e**) and (**f**) Scatter plot of experimental vs predicted pKi values for non-cyclic urea molecules using PLS derived from non-cyclic urea set

Fig. 5 CoMFA steric standard deviation (S.D.* coefficient) contour maps illustrating steric features in combination with compound (**a**) 52 and (**b**) NCU16. Green contours show favorable bulky group substitution at that point while yellow regions show disfavorable bulky group for activity
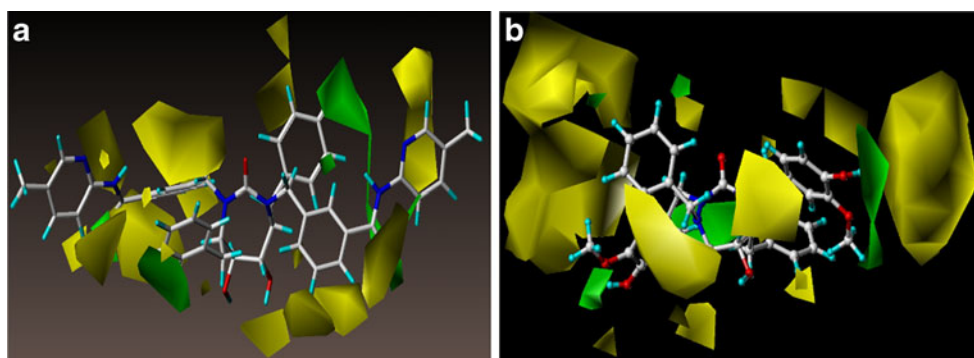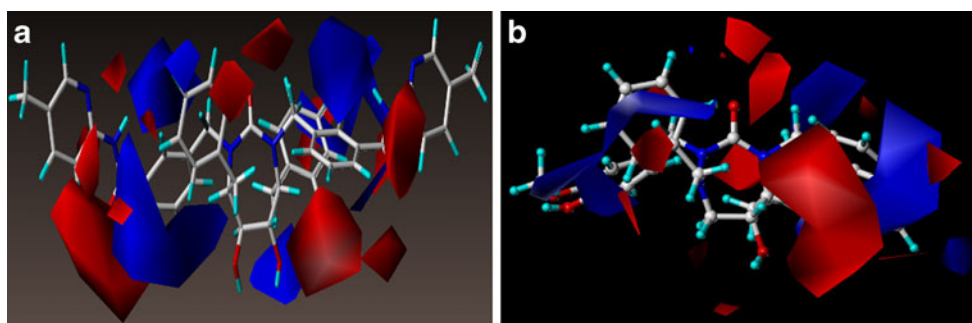
Fig. 6 CoMFA electrostatic standard deviation (S.D.* coefficient) contour maps illustrating electrostatic features in combination with compound (**a**) 52 and (**b**) NCU16. Red contours indicate negative charge favoring activity, whereas blue contours indicate positive charge favoring activity

showing that the two contours for cyclic and non-cyclic urea molecules are similar. This depicts that in cyclic urea molecule substitution on the benzyl ring at R^2 position will increase the activity, in non-cyclic urea molecules increase in bulkiness at R^2 position for triazalinones, diaza sulfoxide, substitution at 3rd and 7th position of coumarin moiety and 5,6 position of pyranones will increase the enzyme inhibitory activity.

Figure 6a, b shows the CoMFA electrostatic contour maps for cyclic urea and non-cyclic urea molecules respectively. The blue and red contours depict the positions where positively charged groups and negatively charged groups would be beneficial for inhibitory activity. In both the contours a red region is seen near the carbonyl and hydroxyl groups of the cyclic urea scaffold and non-cyclic urea molecules, suggesting an electron withdrawing group will be preferred at this position.

The huge blue region around the two amino groups suggests, electron donating group will be beneficial. The red contour region around the side chain atoms of R^1 nearer to the carbonyl group also suggests that an electro negative group will be a potential substituent. In non-cyclic urea molecule there is a disparity in the contour's suggesting a flip in the molecules toward the right where the contours are more concentrated. This suggests that carbonyl group and the hydroxyl group are the pivotal substituents for the

increase of inhibitory activity. The CoMSIA steric and electrostatic field contour maps were almost similar to the corresponding CoMFA contour maps

Hydrophobic contour maps

The hydrophobic fields are presented in Fig. 7, yellow and white contours highlight areas where hydrophobic and hydrophilic groups are preferred respectively. The hydrophobic contour shows the presence of large yellow region near benzyl ring of R^2 substitution in cyclic urea scaffold for the compound 52, yellow contour are scattered over the benzyl rings of R^2 substituents on the non-cyclic urea molecule NCU16, indicates that the groups with hydrophobic characters are preferred at these positions. White hydrophilic favored contour is observed on the amide group of the R^1 position of the compound 52, suggesting group having hydrogen bond forming ability at these positions will be beneficial for protein binding, which is evident form the docking studies as shown in Fig. 8 where the amide group is interacting with Asp 30 and Gly 48 of protein active site. In non-cyclic urea molecules the hydrophilic contour is observed on hydroxyl and methoxy groups of the template NCU16 at R^1 position, this indicates similarity in contours obtained from the two different QSAR model generated from a different set of molecules with diverse scaffold.

Fig. 7 CoMSIA hydrophobic standard deviation (S.D.* coefficient) contour maps illustrating hydrophobic features in combination with compound (**a**) 52 and (**b**) NCU16. Yellow contours indicate hydrophobic group favored region , white contours indicate hydrophilic group favored region
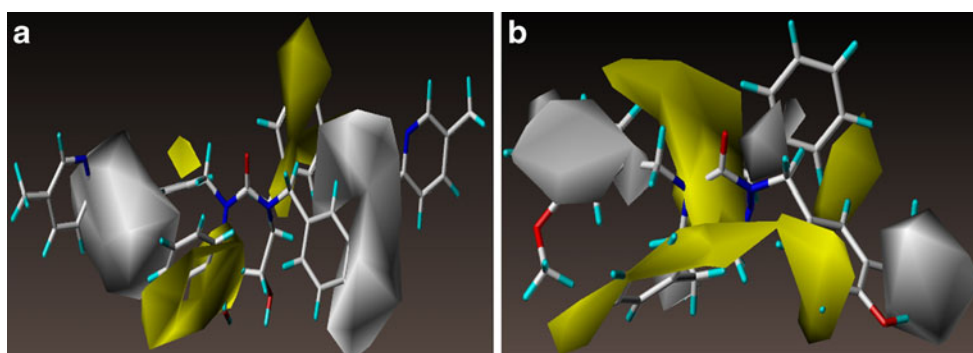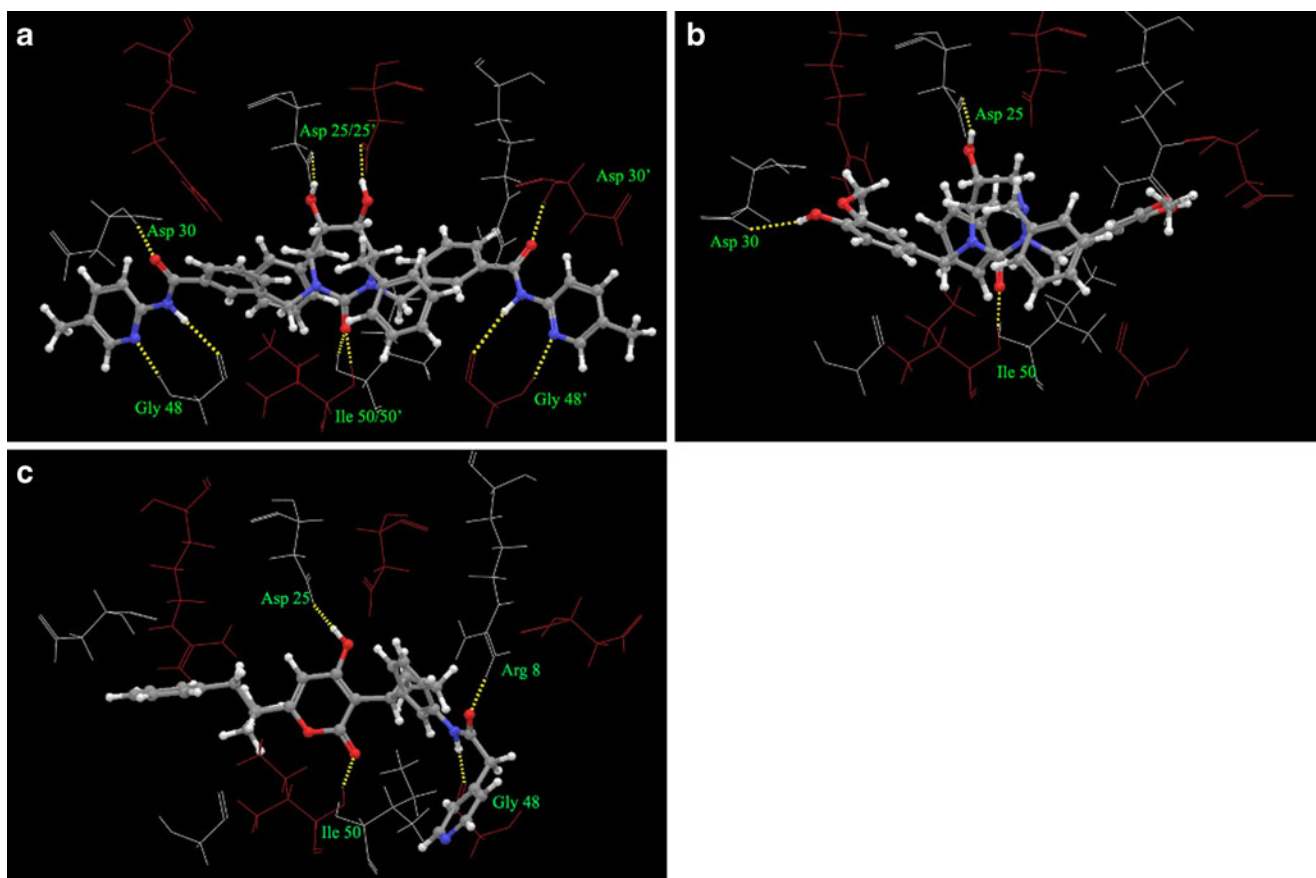
Fig. 8 Docked pose of molecules in the protein active site showing hydrogen bond interaction with the active site amino acids (**a**) compound 52, (**b**) compound NCU16 and (**c**) compound NCU 8

Fig. 9 CoMSIA H-bond acceptor standard deviation (S.D.* coefficient) contour maps illustrating H-bond acceptor features in combination with compound (**a**) 52 and (**b**) NCU16. Magenta contours indicate H-bond acceptor group favored region, red contours indicate H-bond acceptor group disfavored region. CoMSIA H-bond donor standard deviation (S.D.* coefficient) contour maps illustrating H-bond donor features in combination with compound (**c**) 52 and (**d**) NCU16. Cyan contours indicate H-bond donor group favored region, purple contours indicate H-bond donor group disfavored region
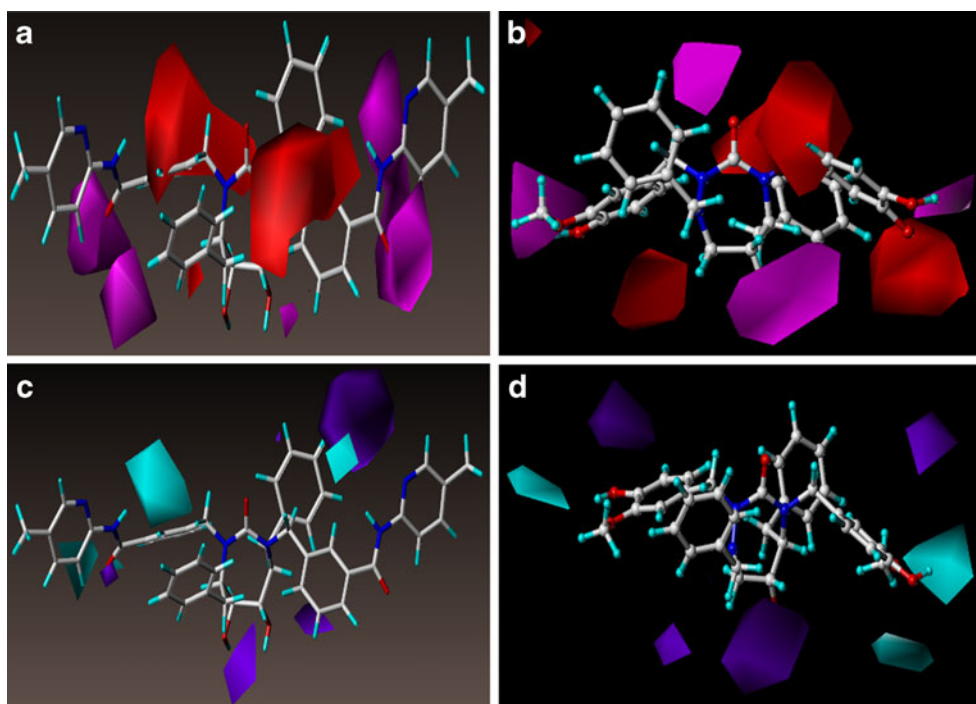
Acceptor and donor contour maps

The hydrogen bond acceptor and donor field contour maps of CoMSIA is shown in Fig. 9a-d using the same templates of cyclic and non-cyclic urea. The magenta and red contours represent favorable and unfavorable hydrogen bond acceptor groups respectively, cyan and purple contours represent favorable and disfavorable hydrogen bond donor groups respectively.

The magenta contour near the carbonyl of the amide group on substituent at R^1 position reveals that hydrogen bond acceptor group may increase the inhibitory activity. This is in agreement with the donor contour where a small purple region for hydrogen bond donor disfavored is seen at the same position. This is in accordance with the docking results as shown in Fig. 8 where the carbonyl of the amide group on substituent at R^1 position in compound 52 shows hydrogen bond interaction with Asp 30 in protein active site. There is a big acceptor disfavored red contour near NH of amide group and surrounding the carbonyl of the cyclic scaffold, this is similar to the hydrogen bond donor contour were a cyan contour is seen at the same region which is favored for hydrogen bond donor groups. The NH group at this position shows a hydrogen bond interaction with Gly 48. Hydrogen bond acceptor and donor contours for non-cyclic urea obtained are similar to the cyclic urea showing same contour regions. The contour map analysis indeed show that the derived 3D QSAR model from the two set of molecules having diverse structural motifs which were aligned based on multiple receptor conformation docking and clustering as a prerequisite have consensus.

The present reported approach of multiple receptor conformation docking and clustering for obtaining receptor based conformation of ligands for CoMFA and CoMSIA analysis, is applicable for ligands having specific protein targets with at least eight to ten experimental three dimensional structures, but not for the targets having less or none. For such cases a homology modeling can be performed and different receptor conformations can be obtained for docking.

Conclusions

A new approach of multiple receptor conformation docking and clustering was employed to obtain a flexible receptor based alignment of molecules having diverse motifs, which could not be aligned based on common substructure that is a prerequisite for CoMFA and CoMSIA analysis. The statistical results and contours obtained invariably show that the method employed is having an adequate accuracy, and provides valuable information for the structural requirements for improving inhibitory activity of the molecules. This advanced and novel approach is appropriate for receptor based alignment for molecules having varied structural motifs, recommending an increase in accuracy of 3D QSAR predictions for considering diverse scaffolds while screening and designing molecules for specific targets.

References

1. Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, Mann D, Sidhu GD, Stahl RE, Zolla-Pazner S, Leibowitch J, Popovic M (1983) Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science 220:865–867
2. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vézinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220:868–871
3. Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RAF, Scolnick EM, Sigal IS (1988) Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci USA 85:4686–4690
4. Tomasselli AG, Howe WJ, Sawyer TK, Wlodawer A, Heinrikson RL (1991) The complexities of AIDS: An assessment of the HIV protease as a Therapeutic Target. Chim Oggi 9:6–27
5. Huff JR (1991) HIV Protease A Novel chemotherapeutic Target for AIDS. J Med Chem 34:2305–2314
6. Norbeck DW, Kempf DJ (1991) HIV Protease Inhibitors. Annu Rep Med Chem 26:141–160
7. Debouck C (1992) The HIV-1 protease as a therapeutic target for AIDS. AIDS Res Hum Retroviruses 8:153–164
8. Katz RA, Skalka AM (1994) The Retroviral Enzymes. Annu Rev Biochem 63:133–173
9. Lam PYS, Jadhav PK, Eyermann CJ, Hodge CN, Ru Y, Bacheler LT, Meek JL, Otto MJ, Rayner MM, Wong YN, Chang CH, Weber PC, Sharpe TR, Jackson DA, Erickson-Viitanen S (1994) Rational design of potent bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. Science 263:380–384
10. Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967
11. 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
12. Kubinyi H, Folkers G, Martin YC (eds) (1998) 3D-QSAR in drug design, vol 2. Kluwer,Dordrecht, the Netherlands
13. Veerapandian P (ed) (1997) Structure-Based Drug Design. Dekker, New York
14. Cho SJ, Garsia MLS, Bier J, Tropsha A (1996) Structure-Based Alignment and Comparative Molecular Field Analysis of Acetylcholinesterase Inhibitors. J Med Chem 39:5064–5071
15. Sippl W, Contreras JM, Parrot I, Rival YM, Wermuth CG (2001) Structure-based 3D QSAR and design of novel acetylcholinesterase inhibitors. J Comput-Aided Mol Des 15:395–410

16. Sippl W (2000) Receptor-based 3D QSAR analysis of estrogen receptor ligands – merging the accuracy of receptor-based alignments with the computational efficiency of ligand-based methods. J Comput-Aided Mol Des 14:559–572

17. McGovern DL, Mosier PD, Roth BL, Westkaemper RBK (2010) CoMFA analyses of C-2 position Salvinorin A analogs at the kappa-opioid receptor provides insights into epimer selectivity. J Mol Graph Model 28:612–625

18. Muddassar M, Pasha FA, Chung HW, Yoo KH, Oh CH, Cho SJ (2008) Receptor guided 3D-QSAR: a yseful approach for designing of IGF-1R inhibitors. J Biomed Biotechnol, vol. 2008, Article ID 837653. doi:10.1155/2008/837653

19. Jojart B, Marki A (2007) Receptor-based QSAR studies of non-peptide human oxytocin receptor antagonists. J Mol Graph Model 25:711–720

20. Carlson HA, Masukawa KM, McCammon JA (1999) Method for including the dynamic fluctuations of a protein in computer-aided drug design. J Phys Chem 103:10213–10219

21. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, Lins RD, Briggs JM, McCammon JA (2000) Developing a dynamic pharmacophore model for HIV-1 integrase. J Med Chem 43:2100–2114

22. Kanth SS, Vijjulatha M (2008) Tetrahydroxy cyclic urea-potent inhibitor for HIV-1 protease wild type and mutant type - a computational design. E-J Chem 5:584–592

23. Sham HL, Zhao C, Stewart KD, Betebenner DA, Lin S, Park CH, Kong XP, Rosenbrook W Jr, Herrin T, Madigan D, Vasavanonda S, Lyons N, Molla A, Saldivar A, Marsh KC, McDonald E, Wideburg NE, Denissen JF, Robins T, Kempf DJ, Plattner JJ, Norbeck DW (1996) A novel, picomolar inhibitor of human immunodeficiency virus type 1 protease. J Med Chem 39:392–397

24. Ala PJ, Huston EE, Klabe RM, Jadhav PK, Lam PY, Chang CH (1998) Counteracting HIV-1 protease drug resistance: structural analysis of mutant proteases complexed with XV638 and SD146, cyclic urea amides with broad specificities. Biochemistry 37:15042–15049

25. Backbro K, Lowgren S, Osterlund K, Atepo J, Unge T, Hulten J, Bonham NM, Schaal W, Karlen A, Hallberg A (1997) Unexpected binding mode of a cyclic sulfamide HIV-1 protease inhibitor. J Med Chem 40:898–902

26. Huang PP, Randolph JT, Klein LL, Vasavanonda S, Dekhtyar T, Stoll VS, Kempf DJ (2004) Synthesis and antiviral activity of P1' arylsulfonamide azacyclic urea HIV protease inhibitors. Bioorg Med Chem Lett 14:4075–4078

27. Jadhav PK, Ala P, Woerner FJ, Chang CH, Garber SS, Anton ED, Bacheler LT (1997) Cyclic urea amides: HIV-1 protease inhibitors with low nanomolar potency against both wild type and protease inhibitor resistant mutants of HIV. J Med Chem 40:181–191

28. Lam PY, Ru Y, Jadhav PK, Aldrich PE, DeLucca GV, Eyermann CJ, Chang CH, Emmett G, Holler ER, Daneker WF, Li L, Confalone PN, McHugh RJ, Han Q, Li R, Markwalder JA, Seitz SP, Sharpe TR, Bacheler LT, Rayner MM, Klabe RM, Shum L, Winslow DL, Kornhauser DM, Hodge CN (1996) Cyclic HIV protease inhibitors: synthesis, conformational analysis, P_2/P_2' structure-activity relationship, and molecular recognition of cyclic ureas. J Med Chem 39:3514–3525

29. Jadhav PK, Woerner FJ, Lam PY, Hodge CN, Eyermann CJ, Man HW, Daneker WF, Bacheler LT, Rayner MM, Meek JL, Erickson-Viitanen S, Jackson DA, Calabrese JC, Schadt M, Chang CH (1998) Nonpeptide cyclic cyanoguanidines as HIV-1 protease inhibitors: synthesis, structure-activity relationships, and X-ray crystal structure studies. J Med Chem 41:1446–1455

30. Hodge CN, Aldrich PE, Bacheler LT, Chang CH, Eyermann CJ, Garber S, Grubb M, Jackson DA, Jadhav PK, Korant B, Lam PY, Maurin MB, Meek JL, Otto MJ, Rayner MM, Reid C, Sharpe TR, Shum L, Winslow DL, Erickson-Viitanen S (1996) Improved cyclic urea inhibitors of the HIV-1 protease: synthesis, potency, resistance profile, human pharmacokinetics and X-ray crystal structure of DMP 450. Chem Biol 3:301–314

31. Schaal W, Karlsson A, Ahlsen G, Lindberg J, Andersson HO, Danielson UH, Classon B, Unge T, Samuelsson B, Hulten J, Hallberg A, Karlen A (2001) Synthesis and comparative molecular field analysis (CoMFA) of symmetric and nonsymmetric cyclic sulfamide HIV-1 protease inhibitors. J Med Chem 44:155–169

32. Yamazaki T, Hinck AP, Wang YX, Nicholson LK, Torchia DA, Wingfield P, Stahl SJ, Kaufman JD, Chang CH, Domaille PJ, Lam PY (1996) Three-dimensional solution structure of the HIV-1 protease complexed with DMP323, a novel cyclic urea-type inhibitor, determined by nuclear magnetic resonance spectroscopy. Protein Sci 5:495–506

33. Schrödinger LLC (2005) Glide, Version 4.0. New York, NY

34. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw ED, 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

35. Nugiel DA, Jacobs K, Worley T, Patel M, Kaltenbach RF III, Meyer DT, Jadhav PK, de Lucca GV, Smyser TE, Klabe RM, Bacheler LT, Rayner MM, Seitz SP (1996) Preparation and structure-activity relationship of novel P1/P1'-substituted cyclic urea-based human immunodeficiency virus type-1 protease inhibitors. J Med Chem 39:2156–2169

36. Patel M, Kaltenbach RF III, Nugie DA, Mchugh RJ Jr, Jadhav PK, Bacheler LT, Cordova BC, Klabe RM, Erickson-Viitanen S, Garber SS, Ried C, Sitz SP (1998) The synthesis of symmetircal and unsymmetrical P_1/P_1' Cyclic ureas as HIV protease inhibitors. Bioorg Med Chem Lett 8:1077–1082

37. Patel M, Bacheler LT, Rayner MM, Cordova BC, Klabe RM, Erickson Viitanen S, Sitz SP (1998) The synthesis and evaluation of cyclic ureas as HIV protease inhibitors: modification of the P_1/P_1' groups. Bioorg Med Chem Lett 8:823–828

38. Rodgers JD, Johnson BL, Wang H, Erickson-Viitanen S, Klabe RM, Bacheler LT, Cordova BC, Chang CH (1998) Potent cyclic urea HIV protease inhibitors with 3-aminoindazole P_2/P_2'groups. Bioorg Med Chem Lett 8:715–720

39. Ax A, Schaal W, Vrang L, Samuelsson B, Hallberg A, Karlén A (2005) Cyclic sulfamide HIV-1 protease inhibitors, with sidechains spanning from P2/P2′ to P1/P1′. Bioorg Med Chem 13:755–764

40. Hultén J, Andersson HO, Schaal W, Danielson HU, Classon B, Kvarnström I, Karlén A, Unge T, Samuelsson B, Hallberg A (1999) Inhibitors of the C_2-symmetric HIV-1 protease: nonsymmetric binding of a symmetric cyclic sulfamide with ketoxime groups in the P2/P2′ side chains. J Med Chem 42:4054–4061

41. Sybyl version 6.9 (1999) Tripos Associates, St. Louis (MO)

42. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges. Tetrahedron 36:3219–3228

43. Wheelock CE, Nakagawa Y, Harada T, Oikawa N, Akamatsu M, Smagghe G, Stefanou D, Iatrou K, Swevers L (2006) High-throughput screening of ecdysone agonists using a reporter gene assay followed by 3-D QSAR analysis of the molting hormonal activity. Bioorg Med Chem 14:1143–1159

ORIGINAL PAPER

An important factor in relation to shock-induced chemistry: resonance energy

Bisheng Tan · Rufang Peng · Xinping Long ·
Hongbo Li · Bo Jin · Shijin Chu

Abstract With density function theory BLYP/DNP method, together with homodesmotic reactions and isodesmic reactions, we calculated the resonance energies of some explosives, including eight nitro compounds which contains benzene rings, three nitro compounds which contains azaheterocycles (2,4-dinitroimidazole (2,4-DNI), 2,6-diamino-3,5-dinitropyrazine-1-oxide (LLM-105) and 2,4,6-trinitro-1,3,5-triazine) and one nitrogen-rich energetic compound of 3,3'-azobis(6-amino-s-tetrazine) (DAAT). The results indicate that their resonance energies are in relation to their shock sensitivity which measuring their threshold pressures of initiation, that is, the lower the resonance energy is, the higher the shock sensitivity of the explosive behaves. And this measuring method according to resonance energy is based on the global property of the molecule instead of the local one, such as one nitro group in the molecule. It is meaningful to calculate resonance energies of these kind of compounds quickly and accurately because resonance struc-
tures exist widely in these organic compounds and resonance energies may play a significant role in determining their shock sensitivity, and it is helpful in the rational design or synthesis of high energy and insensitive materials.

Introduction

Shock wave initiating detonation of an explosion is a complicated physical and chemical process. The vibrational energy up-pumping model [1] suggests that the shock wave produces a bath of excited phonons absorbed by the lowest vibrational modes of molecules that make up the crystal. Increased phonon absorption and intramolecular vibrational energy redistribution lead to excitation of higher frequency modes, eventually lead to a transition state, which will result in chemical bond breakage and subsequent chemical reactions. The rate of phonon-to-vibron energy transfer is related to the sensitivity in common explosives [2]. The multidimensional reactive flow models of shock initiation and detonation of solid explosives were developed by Tarver [3]. These models started from the multiphonon up-pumping, early unimolecular reactions open most channels of the available energy into excited vibrational states of intermediate product species. The intermediate products transfer some of their vibrational energy back into the transition states and accelerate the overall reaction rates. As the decomposition progresses, the high vibrational excited diatomic and triatomic molecules formed during exothermic chain reactions achieve rapid vibrational equilibrium by "super-

B. Tan (✉) · R. Peng · H. Li · B. Jin · S. Chu
State Key Laboratory Cultivation Base
for Nonmetal Composites and Functional Materials,
Southwest University of Science and Technology,
Mianyang, People's Republic of China 621010
e-mail: cqtbs@163.com

B. Tan · X. Long
School of Mechano-Electronic Engineering,
Beijing Institute of Technology,
Beijing, People's Republic of China 100081

X. Long (✉)
China Academy of Engineering Physics (CAEP),
Mianyang, People's Republic of China 621900
e-mail: longxinping@vip.sina.com

collisions" by which is transfered large amounts of vibrational energy between these molecules. The induction period for the initial endothermic bond breaking reaction can be calculated with the high-pressure, high-temperature transition state theory.

Raman spectrum is an important measure of studying the relationship between the shock sensitivities and the molecular structures. Like McNesby and Coffey [2] in the case of impact sensitivity, Koshi [4] had also used Raman spectra in the characterization of the relationships between the shock/impact sensitivities and the molecular structure. With the development of time-resolved optical spectroscopy, some researchers began to study initiation reaction caused by shock in view of it. Patterson [5] used this method to examine chemical decomposition of RDX crystals shocked along the [1 1 1] orientation to peak stresses between 7 and 20 GPa; McGrane [6] applied ultrafast laser shock generation methods and broad-band infrared reflection absorption spectroscopy to study the shock-induced chemistry in a condensed-phase energetic material, in which fission occurs on a time scale of tens of picoseconds and it involves the nitro group as a primary point. And femtosecond broad-band infrared reflection absorption spectroscopy was used to monitor films of the energetic polymer poly (vinyl nitrate) during shock loading and rarefaction.

Nanoshocks recommended by Dlott [7] are tiny but powerful laser-driven shock waves that can be used to produce large-amplitude compression in molecular materials on the picosecond time scale. When coupled with ultrafast molecular spectroscopy, the molecular response to nanoshocks can be probed in detail. Now nanoshocks are used to study shock-induced chemical reactions of energetic materials.

It is very helpful to study initiation by shock for explosives at molecular level to reveal the mechanism of initiation, but so far, the knowledge is still not convincing and ambiguous to illustrate mechanism of initiation. Investigating shock sensitivity at the macro-scale levels is still very popular and it can be substituted by microcosmic means. Just the same as the conventional macroscopic titrations tests in modern analytical science remain irreplaceable even a lot of high resolution analytical methods spring up, so the data of shock sensitivity obtained by conventional small-scale gap test is therefore precious. Shock sensitivity is a kind of measuring method to probe the critical pressure of initiation obtained by small-scale gap test or wedge test. A drop hammer test is still popular because of its simplicity, but it is not like small-scale gap test which has stricter definition. Some scholars have attempted to find the relationship between the shock sensitivity of the limited explosive compounds and their

molecular structures, and they had made great progress and reached important fruits. Politzer and Murray [8] predicted the impact sensitivities of some explosives in terms of molecular electrostatic potential; Rice [9] calculated the bond dissociation energies of the weakest $C-NO_2$ bonds in some nitro-compounds by means of density functional theory. Fried [10] argued that bond dissociation energy alone was not enough to capture high explosive sensitivity; Rice [11] also investigated the relationship between the impact sensitivities of energetic materials and their charge distributions; Owens [12] studied the energy barriers of bond breaking of some energetic materials; Zeman [13] demonstrated the relationship between the detonation velocity of explosives and their nitro electrostatic charges; Zeman [14] also proposed a viewpoint termed as "trinitro-toluene mechanism" to explain the decomposition paths of TNT, TATB, et al. under the condition of shock; Zhang [15] studied the Mulliken charges of nitro group of some nitro compounds by DFT/BLYP/DNP and found the relationship between the impact sensitivities of these compound and their Mulliken charges of nitro groups.

At present, too much emphasis is focused on the local positions of a molecule (such as nitro group) rather than on the global ones when the relationship between the shock sensitivity and the molecular structures need to be disclosed and the data mostly applied is the heights of hammer, H_{50}. Recently, our research group also found that bond dissociation energy of $X-NO_2$ (X = C, N, or O) and Mulliken charges of nitro groups are important factors influencing shock sensitivity [16] of nitro compounds. In this work we start from a new angle, resonance energy, a global property of a molecule rather than a local one (such as the property of a nitro group) to study how the resonance energy influences the shock sensitivity in terms of the calculations of eight nitro compounds which contains benzene rings, three nitro compounds which contain azaheterocycles (2,4-DNI, LLM-105 and 2,4,6-trinitro-1,3,5-triazine) and one nitrogen-rich energetic compound (DAAT). Resonance energy is a kind of stabilization energy owing to delocalization of electrons, and the aromaticity of a benzene ring and the other heterocycles are strongly in relation to resonance energy.

Calculational details

All of the electronic structure calculations were performed using the DMol3 [17, 18] numerical-based density-functional computer software implemented in the Materials Studio Modeling 3.0 package [19] distributed by Accelrys, Inc. Geometrical optimizations were obtained by using the BLYP [20, 21] general-gradient potential approximation in

conjunction with the double-numerical plus polarization basis set which was denoted as DNP. A series of homodesmotic reaction equations [22, 23] were designed to calculate resonance energy. Energy difference in a homodesmotic reaction equation denotes the resonance energy. The bond types and valence types in a homodesmotic reaction equation are kept the same and this treatment can decrease the systematic error of the calculation to a significant extent. When the calculation value (130.6 kJ mol^{-1}) according to Eq. 1 is scaled by the factor of 1.07, it is in agreement with the value (139.5 kJ mol^{-1}) published [23]. Accordingly, Eqs. 2~9 were designed to calculate the resonance energies of *N*-methyl-*N*-2,4,6-tetranitroaniline (TETRYL), 1,3,5-trinitrobenzene (TNB), 2,4,6-trinitrotoluene (TNT), 2,4,6-trinitrobenzene-1,3,5-triamine (TATB), 2,4,6-trinitrobenzene-1,3-diamine (DATB), 2,4,6-trinitroaniline (MATB), (E)-bis(2,4,6-trinitrophenyl)diazene (HNAB), and 1,3,5-trinitro-2-[(E)-2-(2,4,6-trinitrophenyl) vinyl]benzene (HNS), respectively.

(1)

(2)

(3)

(4)

(5)

(6)

(7)

(8)

(9)
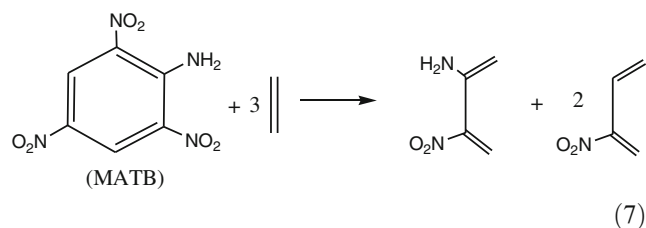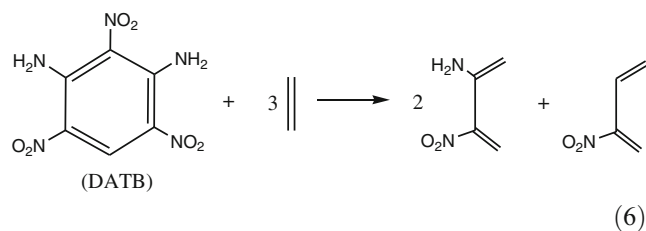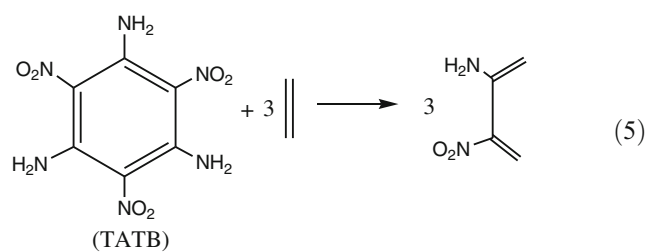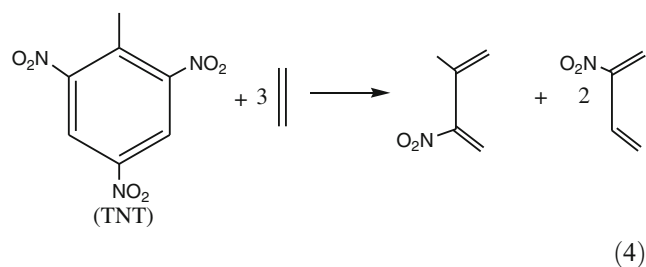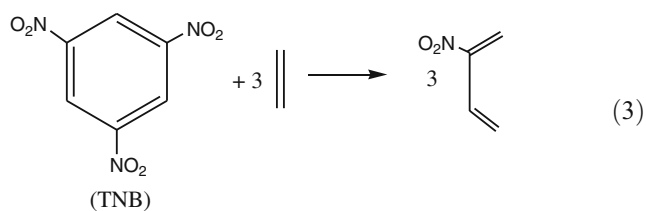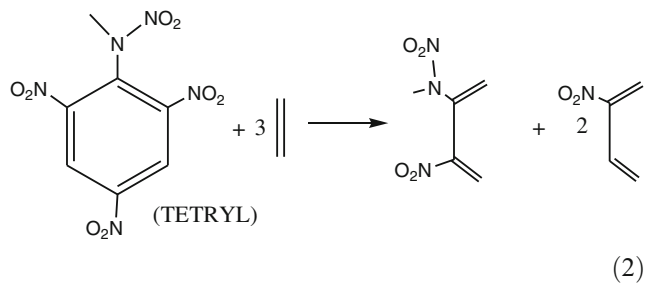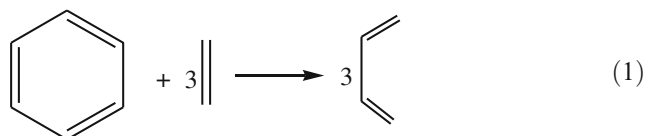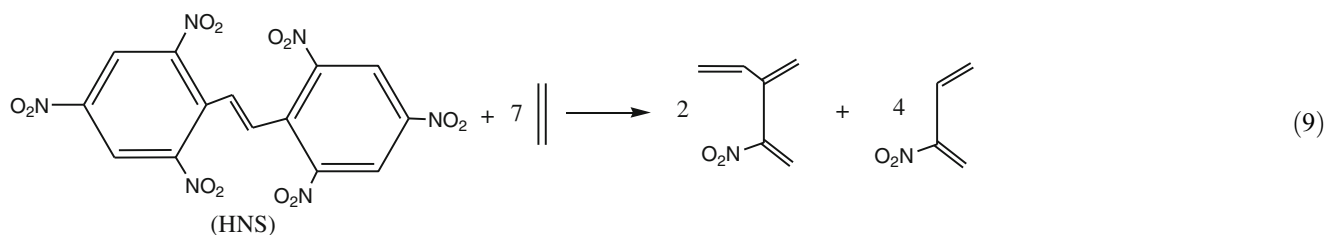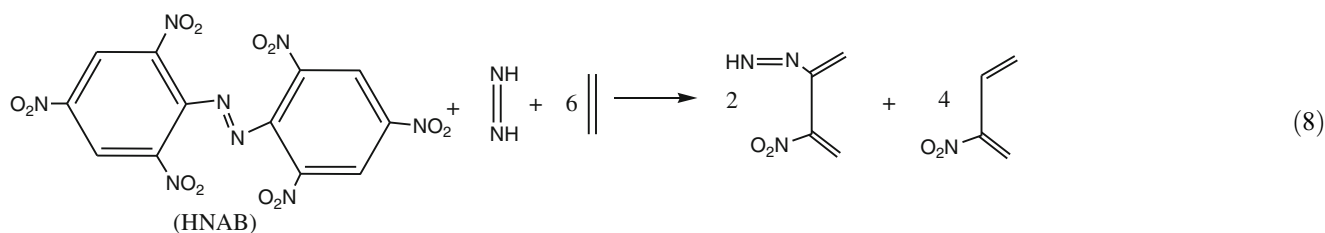
Results and discussions

The "RE" column represents the calculated resonance energies of some explosive compounds in Table 1. From TETRYL to TATB, the resonance energies increase gradually, and their pressures of shock initiation also show the same trend, that is, the lower the resonance energy is, the higher the shock sensitivity the explosive behaves. The forementioned calculation method of resonance energy can be introduced to calculate nitro heterocyclic compounds,

because there is a shortage of shock sensitivity data of these kind of compounds. It is meaningful to measure their shock sensitivity in terms of resonance energy.

Take 2,4-dinitroimidazole (2,4-DNI), 2,6-diamino-3,5-dinitropyrazine-1-oxide (LLM-105), and 2,4,6-trinitro-1,3,5-triazine (TNTA) as examples of nitro heterocyclic compounds to calculate resonance energies. To calculate the resonance energies of 2,4-DNI, LLM-105, and TNTA, the resonance energies of imidazole, pyrazine and triazine must be determined because the latters are the formers' nitro derivatives, respectively. So three equations were firstly designed according to Huang's suggestion [24], see Eqs. 10~12.

Table 1 Pressure (kbar) of shock initiation of some explosive compounds and their calculated resonance energies

Names	RE (kJ/mol)	$P_{90\%, \text{TMD}}$	$P_{95\%, \text{TMD}}$	$P_{98\%, \text{TMD}}$
TETRYL	39.5	10.64	15.14	19.42
HNAB	95.8	12.77	18.11	22.48
TNT	75.2	17.72	25.65	33.35
TNB	98.0	14.96	27.28	37.25
HNS	110.8	26.26	30.15	32.90
MATB	166.6	27.91	35.35	41.02
DATB	229.7	46.20	54.22	59.88
TATB	272.9	70.38	121.92	164.86

90%TMD, 95%TMD, 98%TMD" means the compressed explosive cylinder's density is 90%, 95%, and 98% of its theoretical maximum density, respectively Ref. [16].

(10)

(11)

$$3 \text{(hexahydrotriazine)} \longrightarrow \text{(triazine)} + 2 \text{(hexahydropyrimidine)} \tag{12}$$

The calculated values in terms of Eqs. 10~12 are lower than the published values by a huge difference, so other equations should be designed. Homodesmotic reaction Eqs. 13 and 14 (similar to Eqs. 1 to 9) are conceived to evaluate resonance energy.

$$\text{(pyrazine)} + \begin{array}{l} 2\,CH_2=NH \\ CH_2=CH_2 \end{array} \longrightarrow 2\,CH_2=CH-N= + HN=CH-CH=NH \tag{13}$$
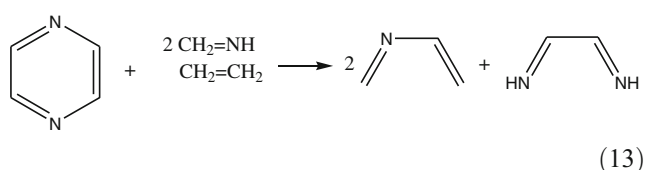
The calculated resonance energy (112.9 kJ mol^{-1}) is lower than in literature [27].

$$\text{(triazine)} + 3\,CH_2=NH \longrightarrow 3\,CH_2=N-CH=NH \tag{14}$$
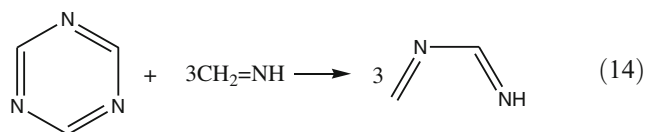
On the contrary, this calculated value (594.5 kJ mol^{-1}) is much higher than the value published [26]. Those homodesmotic reaction equations are not practical, so it should be treated with caution to calculate heterocycles. Taking the four methods suggested by Nyulaszi [25] about five-membered heterocycles into account, Eqs. 15~18 were designed to calculate resonance energy of imidazole. The calculations indicate that the results according to Eqs. 15~17 are evidently lower than the value published [26], only Eq. 18 is close to that in reference. The same way was applied to calculate pyrazine and triazine, it is the same as imidazole. When the results of the calculation are scaled by the factor of 0.6, the calculated resonance energies of imidazole, pyrazine, and triazine are all in good agreement with the values published [26], see Eqs. 19 and 20. This treatment was employed to calculate 2,4-DNI, LLM-105, and TNTA, see Eqs. 21~23. In Eqs. 18 to 23 the bond types are kept the same and the equations are termed as isodesmic reactions [23].

$$\text{(imidazole)} + \begin{array}{l} CH_2=CH_2 \\ CH_2=NH \\ 2\,CH_3NH_2 \end{array} \longrightarrow \text{(vinyl imine)} + \begin{array}{l} CH_2=CHNH_2 \\ H_2NCH=NH \\ CH_3NHCH_3 \end{array} \tag{15}$$

$$\text{(imidazole)} + \begin{array}{l} CH_2=CH-N=CH_2 \\ NH_2-CH=NH \\ NH_2-CH=CH_2 \end{array} \longrightarrow \begin{array}{l} 2\,CH_2=CH-N=CHNH_2 \\ CH_2=CH-NH-CH=NH \end{array} \tag{16}$$

$$\text{(imidazole)} + \begin{array}{l} NH_3 \\ CH_2=CH_2 \\ CH_2=NH \end{array} \longrightarrow \begin{array}{l} CH_2=CH-N=CH_2 \\ CH_2=CHNH_2 \\ NH_2CH=NH \end{array} \tag{17}$$

$$\text{(imidazole)} + \begin{array}{l} 2\,NH_3 \\ 3\,CH_4 \end{array} \longrightarrow \begin{array}{l} CH_2=CH_2 \\ CH_2=NH \\ 3\,CH_3NH_2 \end{array} \tag{18}$$

$$\text{(pyrazine)} + \begin{array}{l} 2\,NH_3 \\ 4\,CH_4 \end{array} \longrightarrow \begin{array}{l} CH_2=CH_2 \\ 2\,CH_2=NH \\ 2\,CH_3NH_2 \\ CH_3CH_3 \end{array} \tag{19}$$

$$\text{(triazine)} + \begin{array}{l} 3\,NH_3 \\ 3\,CH_4 \end{array} \longrightarrow \begin{array}{l} 3\,CH_2=NH \\ 3\,CH_2NH_2 \end{array} \tag{20}$$

$$\text{(2,4-DNI)} + \begin{array}{l} 2\,NH_3 \\ 3\,CH_4 \end{array} \longrightarrow \begin{array}{l} CH_2=CHNO_2 \\ O_2NCH=NH \\ 3\,CH_3NH_2 \end{array} \tag{21}$$

$$(22)$$

$$(23)$$

The corrected resonance energy of imidazole, pyrazine, and triazine is 142.3, 172.4, and 185.7 kJ mol^{-1}, respectively. And the corrected resonance energy of 2,4-DNI, LLM-105, and TNTA is 122.6, 182.4, and 100.7 kJ mol^{-1}, respectively.

The results demonstrate that resonance energy of 2,4-DNI is situated between HNS and MATB, LLM-105 is situated between MATB and DATB, and TNTA is matched with HNS. It can be predicted that their shock sensitivity can be compared with HNS, MATB, and DATB to some extent. This assumption was proved to be rational on the basis of our recent small-scale gap test [27] about 2,4-DNI and LLM-105.

As for nitrogen-rich energetic compound, we take 3,3'-azobis (6-amino-s-tetrazine) (DAAT) as an example and the same treatment as forementioned heterocycles was used to calculate its resonance energy, see Eq. 24.

$$(24)$$
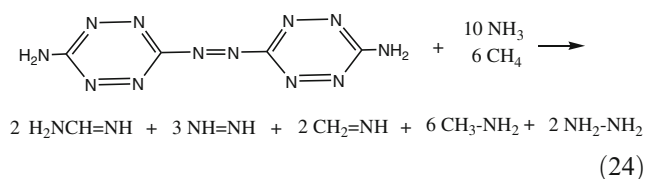
The corrected resonance energy (399.5 kJ mol^{-1}) in terms of Eq. 24 is a very great value and it can be concluded that DAAT is a high nitrogen content and insensitive energetic compound.

It can be speculated that the explosive molecules with resonance structures can disperse the energy caused by shock due to the delocalization movement of π electrons. It is hard to form hot spots which lead to decomposition and therefore their shock sensitivities are usually lower. If a C-NO$_2$ bond or a N-NO$_2$ bond which is the weakest bond in nitro aromatic compounds or nitramine is involved in the resonance structures, its strength will increase because of

the delocalization movement of π electrons. Energy caused by shock can be rapidly transferred through the resonance structures, so the molecule is insensitive to shock. Generally speaking, the higher the resonance energy is, the lower the shock sensitivity is.

With so many factors influencing shock sensitivity, such as particle size, density, crystal form, void, et al. if these influencing factors are kept the same, molecular structures (especially contain resonance structures) may be the key factors which affect shock sensitivity of explosives.

Conclusions

Density function theory BLYP/DNP together with homo-desmotic reactions and isodesmic reactions was employed to calculate resonance energies of eight nitro aromatic compounds, three nitro heterocyclic compounds, and one nitrogen-rich energetic compound. They are ranged as follows: HNB < TETRYL < TNT < HNAB < TNB ≈ TNTA < HNS < 2,4-DNI < MATB < LLM-105 < DATB < TATB < DAAT. It is found that their resonance energies are indeed relevant to their experimental shock sensitivities. So the resonance energies in explosives should not be neglected because they exist widely in energetic materials and may play a significant role in determining their shock sensitivities. Resonance energies are worth studying in the design of high energy and insensitive energetic materials. In addition, strain energy is an important factor which should be paid more attention.

References

1. Tokmakoff A, Fayer MD, Dlott DD (1993) Chemical reaction initiation and hot-spot formation in shocked energetic molecular materials. J Phys Chem 97:1901–1913
2. McNesby KL, Coffey CS (1997) Spectroscopic determination of impact sensitivities of explosives. J Phys Chem B 101:3097–3104
3. Tarver CM (1997) Multiple roles of highly vibrationally excited molecules in the reaction zones of detonation waves. J Phys Chem A 101:4845–4851
4. Wakabayashi K, Matsumura T, Nakayama Y, Yamada E, Koshi M (2007) Temporal change of raman spectra of carbon tetrachloride under laser-driven shock compression. AIP Conf Proc 955:1267–1270
5. Patterson JE, Dreger ZA, Miao M (2008) Shock wave induced decomposition of RDX. J Phys Chem A 112:7374–7382
6. McGrane SD, Moore DS, Funk DJ (2004) Shock induced reaction observed via ultrafast infrared absorption in poly(vinyl nitrate) films. J Phys Chem A 108:9342–9347
7. Dlott DD (2000) Nanoshocks in molecular materials. Acc Chem Res 33:37–45

8. Politzer P, Murray JS (1996) Relationships between dissociation energies and electrostatic potentials of C-NO$_2$ bonds: applications to impact sensitivities. J Mol Struct 376:419–424
9. Rice BM, Sahu S, Owens FJ (2002) Density functional calculations of bond dissociation energies for NO$_2$ scission in some nitroaromatic molecules. J Mol Struct THEOCHEM 583:69–72
10. Fried LE, Manaa MR, Pagoria PF, Simpson RL (2001) Design and synthesis of energetic materials. Annu Rev Mater Res 31:291–321
11. Rice BM, Hare JJ (2002) A quantum mechanical investigation of the relation between impact sensitivity and the charge distribution in energetic molecules. J Phys Chem A 106:1770–1783
12. Owens FJ (1996) Calculation of energy barriers for bond rupture in some energetic molecules. J Mol Struct THEOCHEM 370:11–16
13. Zeman S, Friedl Z, Rohac M (2006) Molecular structure aspects of initiation of some highly thermostable polynitro arenes. Thermochim Acta 451:105–114
14. Varga R, Zeman S (2006) Decomposition of some polynitro arenes initiated by heat and shock part I. 2,4,6-Trinitrotoluene. J Hazard Mater 132:165–170
15. Zhang C, Shu Y, Huang Y, Zhao X, Dong H (2005) Investigation of correlation between impact sensitivities and nitro group charges in nitro compounds. J Phys Chem B 109:8978–8982
16. Storm CB, Stine JR, Kramer JF (1990) In: Bulusu SN (ed) Chemistry and physics of energetic materials. Kluwer Academic Publishers, Dordrecht, pp 605–639
17. Delley B (1990) An all-electron numerical method for solving the local density functional for polyatomic molecules. J Chem Phys 92:508–517
18. Delley B (2000) From molecules to solids with the DMol3 approach. J Chem Phys 113:7756–7764
19. Material Studio 3.0 (2003) Acceryls Inc. San Diego
20. Becke AD (1988) A multicenter numerical integration scheme for polyatomic molecules. J Chem Phys 88:2547–2553
21. 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
22. Wheeler SE, Houk KN, Schleyer PvR, Allen WD (2009) A hierarchy of homodesmotic reactions for thermochemistry. J Am Chem Soc 131:2547–2560
23. Bachrach SM (2002) Aromaticity of annulated benzene, pyridine and phosphabenzene. J Organomet Chem 643–644:39–46
24. Huang YZ, Yang SY, Li XY (2004) An investigation of the aromaticity of transition metal heterocyclic complexes by conventional criteria and indices of aromaticity. J Organomet 689:1050–1056
25. Nyulaszi L, Varnai P, Veszpremi T (1995) About the aromaticity of five-membered heterocycles. J Mol Struct THEOCHEM 358:55–61
26. Bird CW (1997) Heteroaromaticity. 10. The direct calculation of resonance energies of azines and azoles from molecular dimensions. Tetrahedron 53:13111–13118
27. Tan B, Long X, Peng R, Li H, Jin B, Chu S (2011) Insight into shock-induced chemical reaction from the perspective of ring strain and rotation of chemical bonds. J Comput Aided Mol Des, under review

On the electronic properties of two-dimensional honeycomb GaInN and GaAlN alloys: a molecular analysis

Ernesto Chigo Anota · Heriberto Hernández Cocoletzi

Abstract We have performed first principles total energy calculations to investigate the structural and the electronic properties of two-dimensional honeycomb GaAlN and GaInN alloys. Calculations were done using a coronene-like ($C_{24}H_{12}$) cluster and for different numbers of Ga, Al, and In atoms. The exchange and correlation potential energies were treated within the generalized gradient approximation (GGA). The bond length, dipole moment, binding energy, and gap between the HOMO and the LUMO are reported as a function of x. The stability of the structures depends on the site of the substituted atom; for example, when three Ga atoms are substituted, the GaInN alloy becomes unstable. The gap in the GaAlN increases from 3.76 eV (GaN) to 4.51 eV (AlN), and in the GaInN decreases to 2.11 eV. The biggest polarity occurs when eight and four Ga atoms are substituted, for GaAlN and GaInN, respectively.

Keywords Nitride · Coronene · DFT theory · Molecular simulation

Introduction

The obtention of graphene opened investigations into two-dimensional (2D) honeycomb nanostructures [1]. These involve both basic research [2–4] and applications like the doping of graphene for the enhancement of optical properties [5], graphene films for transparent electrodes [6], graphene-based transistors [7], and semiconducting SiC sheets [8], among others. In 2005, Novoselov and coworkers [9] reported the stability of the boron nitride sheet; the group of Şahin [10] then reported the electronic structure of 2D structures of group IV elements and III-V binary compounds; more recently, using a molecular point of view, we have also studied the electronic properties of the III-A nitrides [11]. In order to improve the capabilities and to broaden the applications of these systems, some changes have been proposed; for example, Li and F doping of boron nitride sheet [12], and the insertion of an O atom into graphene [13]. An alternative to these proposals is to consider 2D alloys, as was done with bulk GaN, InN and AlN [14]; it is worth mentioning that these systems can form a set of ternary alloys that can cover the entire visible part of the electromagnetic spectrum. In this work, we present a molecular first principles study analyzing the electronic properties of the graphene-like GaAlN and GaAlN alloys employing a coronene-like model.

Computational details

As in previous studies, the calculations were performed using density functional theory (DFT) [15–18] as implemented in the DMOL^3 code available from Accelrys [19]. We utilized the generalized gradient approximation (GGA) for the exchange-correlation term within the parameterization of Perdew-Burke-Ernzerhof (PBE) [20]. For the GaN, AlN, and InN compounds, we used the configuration $B_{12}N_{12}H_{12}$ with B = Ga, Al, In (Fig. 1). The alloys were studied by considering the $B_{12-x}Al_xN_{12}H_{12}$ (x=1, 2, 3, 4, 8, 12) mesh, which gave rise to the clusters $Ga_{12-x}Al_xN_{12}H_{12}$ and $Ga_{12-x}In_xN_{12}H_{12}$. The average diameter of the circular

E. C. Anota (✉) · H. H. Cocoletzi
Cuerpo Académico Ingeniería en Materiales,
Facultad de Ingeniería Química,
Benemérita Universidad Autónoma de Puebla,
CU San Manuel,
CP 72570 Puebla, Mexico
e-mail: echigoa@yahoo.es
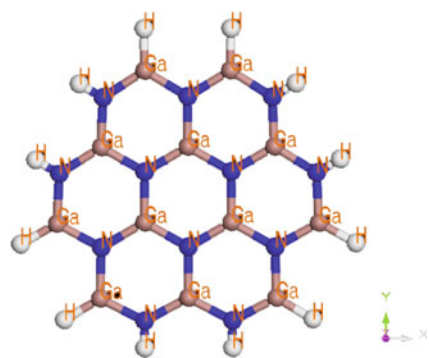
Fig. 1 Cluster model for gallium nitride

sheet is 1.29 nm. The double numeric plus polarization (DNP) all-electron atomic base for the core, in the singlet ground state, was used; this base includes a *p* orbital for the hydrogen atom, and *d* orbitals for nitrogen, aluminum, gallium and indium atoms [19, 21, 22]. The limit for the orbital was 0.50 nm; the convergence for the self consistent field (SCF) cycles was 1.0×10^{-6} Ha. The obtainment of non-negative frequencies was the criterion for structural stability [23]. Additional details on the calculations can be found in [11] and [12], and references therein. We obtain the optimum bond length, the dipole moment, and the vibrational spectrum (stability criterion), as well as binding energy and the difference between HOMO and LUMO (gap).

The cluster size used was validated by calculating the cohesive energy for different clusters; namely, naphthalene type ($Ga_5N_5H_8$), phirene type ($Ga_8N_8H_{10}$), coronene type ($Ga_{12}N_{12}H_{12}$), and the cluster $Ga_{27}N_{27}H_{18}$. We obtained a value of 1.44 a.u./atom for all systems; this means that the size is the correct.

Results and discussion

Several values for x were considered (Fig. 2); isomers other than those used in this work gave similar results. The optimum geometry (all frequencies are positive) of the GaN and AlN sheets is planar, with a Ga–N bond length equal to 1.87 Å and an Al–N bond length equal to 1.82 Å; the Ga–H, and N–H bond lengths are 1.57 and 1.02 Å, respectively. When Ga atoms are substituted by Al atoms, the structure stays stable and planar, and the bond length remains almost unchanged; one new bond, Al–N, appears whose value is 1.81 Å. This is true for the systems $Ga_{11}AlN_{12}H_{12}$ (x=1), $Ga_{10}Al_2N_{12}H_{12}$ (x=2), $Ga_9Al_3N_{12}H_{12}$ (x=3), $Ga_8Al_4N_{12}H_{12}$ (x=4), and $Ga_4Al_8N_{12}H_{12}$ (x=8). For the $Ga_{12-x}In_xN$ alloy we analyzed the same values of x. The In–N bond length is equal to 2.07 Å and the corresponding value for In–H bond length is 1.76 Å. The equivalent notation was also applied to the GaAlN alloy; in this case
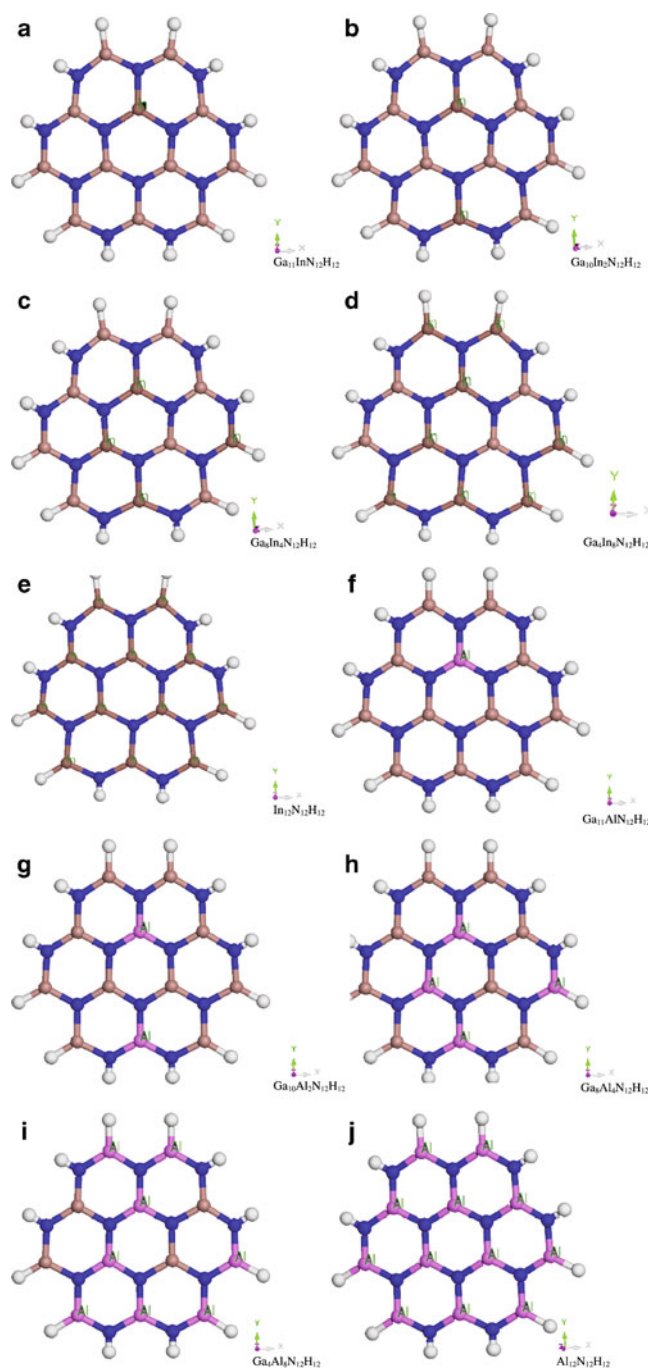
Fig. 2 Cluster models for doped gallium nitride with aluminum and indium

we found that all frequencies are positive for x=1, 2, 4, 8. For x=3 ($Ga_9In_3N_{12}H_{12}$), we found a negative frequency, this means that this structure is unstable and it would be difficult to grow it experimentally; something similar happens with the system $Si_9N_3C_{12}H_{12}$, when the N atom occupies the center of the hexagon [24]; stability depends on the form of doping.

The dipole moment is 4.2×10^{-3} Debye for GaN and 9.1×10^{-3} Debye for AlN. This parameter increases from

x=0 to x=4, then drops to the AlN value, Fig. 3. In the case of the GaInN alloy, this parameter increases from 4.2×10^{-3} Debye to the dipole moment for InN (9.3×10^{-3} Debye). In this case, the increment is higher than in the GaAlN alloy; the maximum value is reached for x=8. The binding energy, as a function of x, is presented in Fig. 4. In the case of GaAlN alloy, this parameter increases as x increases, from GaN (4.62 eV) to AlN (5.51 eV), but for the GaInN alloy the binding energy decreases from GaN to InN (4.01 eV).

When one and four Ga atoms are substituted by Al atoms, the main contribution to the HOMO is due to the p_z nitrogen orbital, up-oriented for the positive lobule and down-oriented for the negative lobule. In the case of two, three and eight substituted atoms, the charge distribution is inverted; the same happens for the AlN sheet. The contribution to the LUMO is due mainly to the p_x and p_z of the nitrogen atoms, and to the p_z boron orbital together with the s hydrogen orbital, in the case of 1, 4, 8, and 12 Al atoms; if the change is by 2 and 3 Al atoms, there are contribution due to s orbital hydrogen atoms and to the p_y orbital nitrogen. When 1, 8 and 12 Ga atoms are substituted by In atoms, we observe a p_z nitrogen (positive lobule up-oriented and negative lobule down-oriented) contribution to the HUMO; and if 2 and 4 Ga atoms are changed, the converse occurs, i.e., the directions of the lobules are inverted. In the case of the LUMO, with the substitution of 1, 4, 8 and 12 Ga atoms, the principal contribution is due to the s and p_z nitrogen atom (positive and negative oriented lobule up oriented down) and to the s orbital hydrogen atom; when two Ga atoms are substituted, the p_z lobules are inverted, mainly in the central hexagon. Figures 5 and 6 summarize all these statements.

For quantum structures, the difference between the HOMO and LUMO (gap) plays a central role in device design. Figure 7 presents the evolution of the gap of
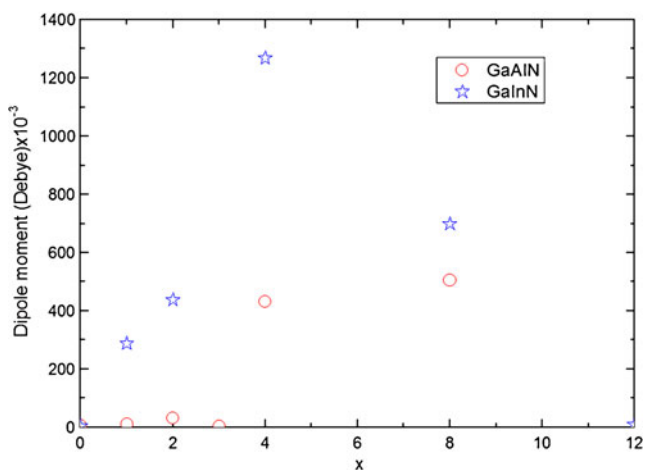
Fig. 4 Binding energy as a function of x for the GaAlN and GaInN alloys

GaAlN sheet alloys as a function of the Al content. For GaN sheet, the gap has a value of 3.67 eV, and when the Ga atoms are substituted by Al atoms this quantity increases monotonically, until the value 4.51 eV for AlN is reached; the gap has increased 18.63%. When the Ga atoms are substituted by In atoms, the gap decreases as the concentration increases, reaching the value 2.11 eV; Fig. 6 shows this behavior. From all this, we can see that, with GaAlN and GaInN alloys, it is possible to cover the entire range of the electromagnetic spectrum from red to ultraviolet. In addition, given the appropriate choice of Al and In content, it would be possible to grow a sheet with a specific gap for a specific optoelectronic application. For example, if one Ga atom is substituted by an In atom, a gap equal to 3.09 eV is obtained; if four In atoms replace the same number of Ga atoms, the corresponding gap is of 2.8 eV. These systems could form the basis of the fabrication of blue emission laser diodes (first case) or green emission devices (second case) as was done with InGaN quantum wells [25, 26]. Recently, Liao et al. [27] grew AlGaN-based deep ultraviolet light emitting diodes; similar devices could be based on the GaAlN sheet choosing x=8 or greater. In general, light sources with short emission wavelengths find applications in high-density optical data storage, biomedical research, water and air purification, and sterilization; such systems can be achieved by varying the number of Al atoms in the GaAlN alloy.

Conclusions

Using first principles total energy calculations and employing the $B_{12}N_{12}H_{12}$ (B = Ga, Al, In) cluster, we have studied

Fig. 3 Dipole moment as a function of x for GaAlN and GaInN alloys

Fig. 5 Isosurfaces of molecular orbitals for gallium nitride doped with Al

HOMO

LUMO

a) with Al; **-5.37 eV**

b) **-1.61 eV**

c) with Al$_2$; **-5.37 eV**

d) **-1.55 eV**

e) with Al$_3$; **-5.37 eV**

f) **-1.42 eV**

g) with Al$_4$; **-5.39 eV**

h) **-1.45 eV**

i) with Al$_8$; **-5.55 eV**

j) **-1.41 eV**

k) with Al$_{12}$; **-5.67 eV**

l) **-1.16 eV**
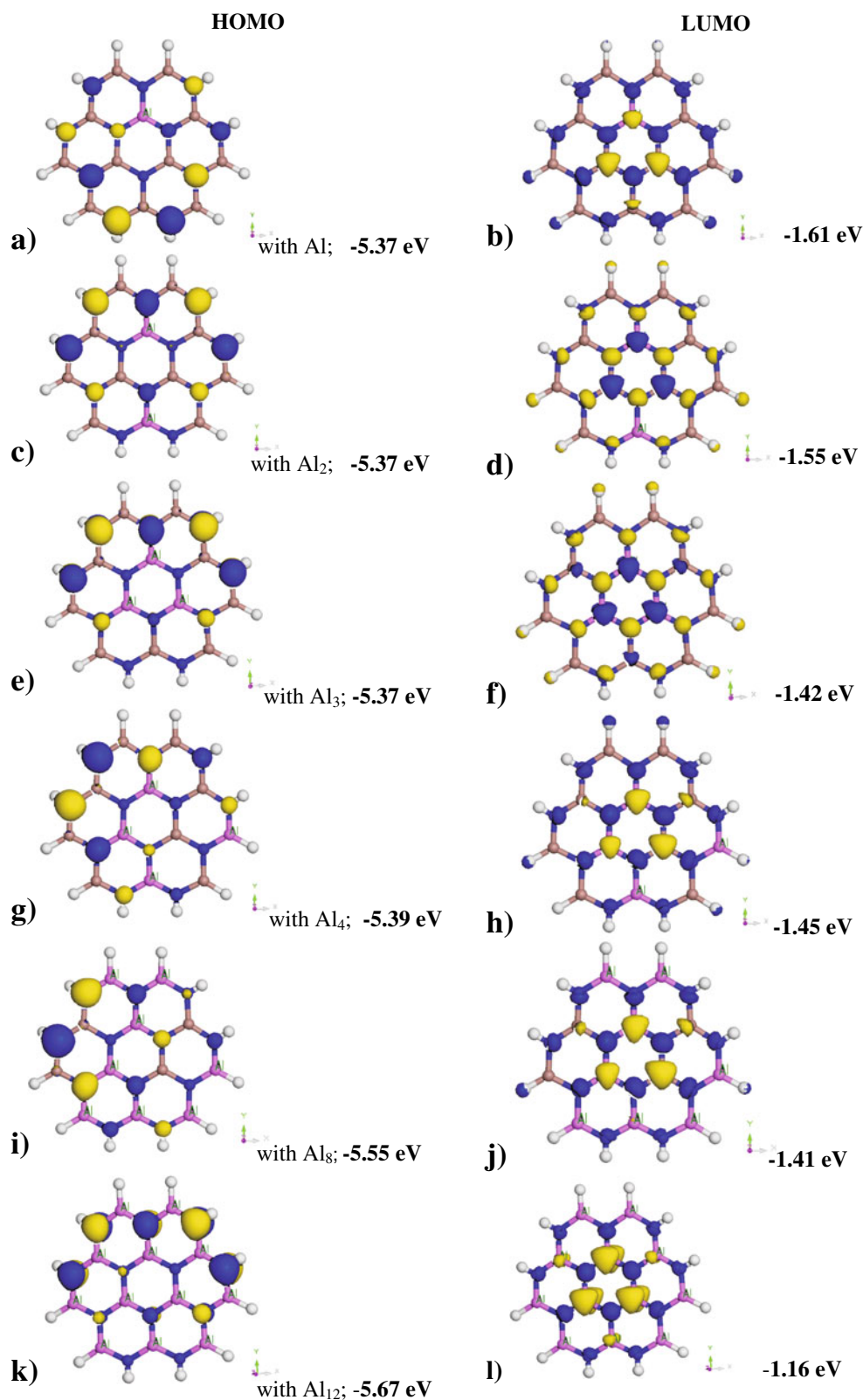
$Ga_{12-x}Al_xN_{12}H_{12}$ and $Ga_{12-x}In_xN_{12}H_{12}$ 2D honeycomb alloys. All the structures studied for the former were stable, and an instability was found for x=3 in the latter. The

higher dipole moment is reached with x=8 for the case of GaInN and x=4 for GaAlN alloy. The binding energy increases as x increases for the GaAlN alloy and decreases

HOMO   **LUMO**

Fig. 6 Isosurfaces of molecular orbitals for gallium nitride doped with In

a)   without In; **-5.37 eV**   b)   **-1.69 eV**

c)   with In; **-5.497 eV**   d)   **-2.28 eV**

e)   with In$_2$; **-5.47 eV**   f)   **-2.38 eV**

g)   with In$_4$; **-5.37 eV**   h)   **-2.57 eV**

i)   with In$_8$; **-5.15 eV**   j)   **-2.71 eV**

k)   with In$_{12}$; **-5.07 eV**   l)   **-2.97 eV**
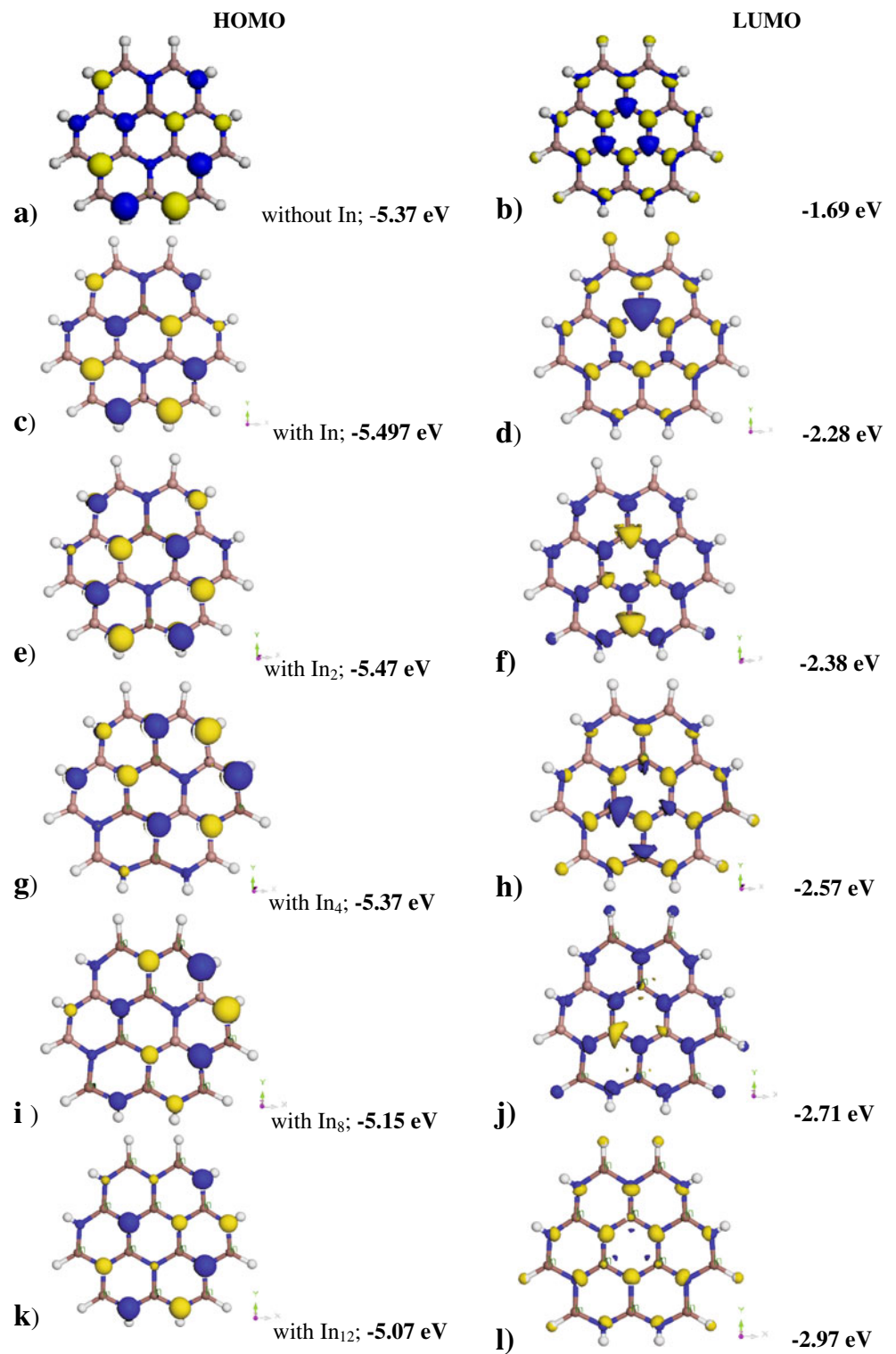
as x decreases for the GaInN alloy. The contribution to the HOMO and to the LUMO is due mainly to p orbitals, and their orientation (up and down) depends on x. The gap between the HOMO and the LUMO increases in going from GaN (4.62 eV) to AlN (5.51 eV), and decreases from GaN to InN (4.01 eV), as in the bulk.
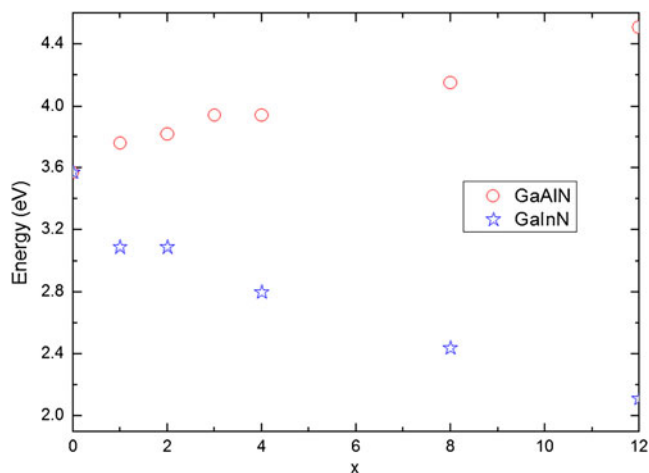
Fig. 7 Gap as a function of x for GaAlN and GaInN alloys

References

1. Novoselov KS, Geim AK, Morozov SV, Jiang D, Zhang Y, Dubonos SV, Grigorieva IV, Firsov AA (2004) Science 306:666–669
2. Zarea M, Sandler N (2007) Phys Rev Lett 99:256804–4
3. Cocco G, Cadelano E, Colombo L (2010) Phys Rev B 81(4):241–412
4. Bistritzer R, MacDonald AH (2010) Phys Rev B 81:245412–245419
5. Bangert U, Bleloch A, Gass MH, Seepujak A, van den Berg J (2010) Phys Rev B 81:245423–11
6. Bae S, Kim H, Lee Y, Xu X, Park JS, Zheng Y, Balakrishnan J, Lei T, Kim HR, Song YII, Kim YJ, Kim KS, Ozyilmaz B, Ahn JH, Hong BH, Iijima S (2010) Nat Nanotechnol 5:574–578
7. Guimarães FSM, Costa AT, Muniz RB, Ferreira MS (2010) Phys Rev B 81:233402–233404
8. Bekaroglu E, Topsakal M, Cahangirov S, Ciraci S (2010) Phys Rev B 81:075433–075439
9. Novoselov KS, Jiang D, Schedin F, Booth TJ, Khotkevich VV, Morozov SV, Geim AK (2005) Proc Natl Acad Sci USA 102:10451–10453
10. Şahin H, Cahangirov S, Topsakal M, Bekaroglu E, Akturk E, Senger RT, Ciraci S (2009) Phys Rev B 80:155453–12
11. Chigo Anota E, Salazar Villanueva M, Hernández Cocoletzi H (2010) Phys Stat Solidi C 7:2252–2254
12. Chigo Anota E, Salazar Villanueva M, Hernández Cocoletzi H (2010) Phys Stat Solidi C 7:2559–2561
13. Dikin DA, Stankovich S, Zimney EJ, Piner RD, Dommett GHB, Evmenenko G, Nguyen ST, Ruoff RS (2007) Nature 448:457–460
14. Hernández Cocoletzi H, Contreras Solorio DA, Arriaga J (2005) Appl Phys A 81:1029–1033
15. Kohn W, Becke AD, Parr RG (1996) J Phys Chem 100:12974–12980
16. Jones RO, Gunnarsson O (1989) Rev Mod Phys 61:689–746
17. Kohn W (1999) Rev Mod Phys 71:1253–1266
18. Chigo Anota E, Rivas Silva JF (2005) Rev Col Fís 37:405–417
19. Delley B (1990) J Chem Phys 92:508–517
20. Perdew JP, Burke K, Ernzerhof M (1996) Phys Rev Lett 77:3865–3868
21. Delley B (1996) J Phys Chem 100:6107–6110
22. Delley B (2000) J Chem Phys 113:7756–7764
23. Foresman JB, Frisch Æ (1996) In: exploring chemistry with electronic structure methods, 2nd edn. Gaussian Inc, Wallingford, CT, p 70
24. Chigo Anota E, Hernández Cocoletzi H, Bautista Hernández A, Sánchez Ramírez JF (2011) J Comput Theor Nanosci 8:637–641
25. Nagahama S, Yanamoto T, Sano M, Mukai T (2002) Phys Status Solidi 194:423–427
26. Li S, Schörmann J, As DJ, Lischka K (2007) Appl Phys Lett 90 (1–3):071903
27. Liao Y, Thomidis C, Kao C, Moustakas TD (2011) Appl Phys Lett 98(1–3):081–110

ORIGINAL PAPER

A DFT study of aminonitroimidazoles

P. Ravi · Girish M. Gore · Surya P. Tewari ·
Arun K. Sikder

Abstract Density functional theory (DFT) calculations at
the B3LYP/aug-cc-pVDZ level were performed to explore
the geometric and electronic structures, band gaps, thermo-
dynamic properties, densities and performances of amino-
nitroimidazoles. The calculated performance properties,
stabilities and sensitivities of the model compounds appear
to be promising compared with those of the known
explosives 2,4-dinitro-1H-imidazole (2,4-DNI), 1-methyl-
2,4,5-trinitroimidazole (MTNI), hexahydro-1,3,5-trinitro-
1,3,5-triazinane (RDX), and octahydro-1,3,5,7-tetranitro-
1,3,5,7-tetraazocane (HMX). The position of the NH_2 or
the number of NO_2 groups on the diazole presumably
determines the structure, heat of formation, stability,
sensitivity, density and performance of the compound.

Keywords Density functional theory · Heat of formation ·
Density · Detonation velocity · Pressure

Introduction

Polynitroimidazoles have recently drawn renewed attention
from explosives chemists due to their highly positive heats
of formation, favorable detonation performances, good

P. Ravi (✉) · S. P. Tewari
Advanced Centre of Research in High Energy Materials
(ACRHEM), University of Hyderabad,
Hyderabad 500 046, India
e-mail: rpiitb@hotmail.com

G. M. Gore · A. K. Sikder
High Energy Materials Research Laboratory (HEMRL),
Pune 411 021, India

A. K. Sikder
e-mail: ak_sikder@yahoo.com

thermal stabilities and insensitivities [1–7]. It is known that
substituting one of the hydrogen atoms of nitroimidazoles
for a NH_2 group increases the stability, heat of formation,
density and performance of such compounds [1]. However,
to our knowledge, there have been no such studies of the
structures and explosive properties of aminonitroimida-
zoles. Therefore, before attempting to synthesize them, it
would be useful to be able to predict their explosion
energies, densities, detonation performances, stabilities and
sensitivities.

The sensitivities or stabilities of energetic compounds
towards impact, shock and heat have been related to the
molecular and electronic structure properties of the com-
pounds [8–13]. Politzer and coworkers [9, 10] correlated
the bond dissociation energies (BDEs) of trigger bonds,
molecular electrostatic potential maxima, and the impact
sensitivities of several nitrocompounds. Pospìšil [14]
showed the relation between crystal volume and impact
sensitivity. Zhang et al. [15, 16] correlated the molecular
electronic structures with the impact sensitivities and
stabilities of known explosives by nitro group charge
analysis. Zeman [17] related the impact and electric spark
sensitivities to the detonation properties, thermal decom-
positions, and ^{13}C and ^{15}N NMR chemical shifts of
polynitrocompounds. The electric spark sensitivity was
related to the squares of the detonation velocity and the
reciprocal temperature, as well as the Piloyan activation
energy and the heat of fusion of the explosive [18]. Zhi et
al. [19] related the lowest unoccupied molecular orbital
energy and the nitro group charge to the electric spark
sensitivity.

The present study aims to aid in the design of amino-
nitroimidazoles for explosives applications by uncovering
the structure–property relationships for these compounds.
Density functional theory (DFT) calculations at the B3LYP/

aug-cc-pVDZ level were therefore carried out to explore the geometries, band gaps, heats of formation, crystal densities, detonation velocities and pressures, impact sensitivities and spark sensitivities of model molecules.

Methods and computational details

Ab initio molecular orbital calculations at the B3LYP/aug-cc-pVDZ level were performed for the model compounds using the Gaussian 03 package [20]. The stationary points for each molecule were positively identified as true local minima on the potential energy surfaces with no imaginary frequencies. Calculations at the B3P86/aug-cc-pVTZ level for the model molecules R30, R32, R65 and R68 were also performed to compare their explosive properties.

Isodesmic reactions have been used to calculate the heats of formation ($\Delta_f H$) of the model compounds. This approach has been shown to be simple and reliable [21]. To minimize errors, the numbers of all kinds of bonds and the imidazole ring, methane, nitromethane and amino-methane were kept constant. The heats of formation for the reference compounds were taken from [22]. $\Delta_f H$ values can be obtained for the model molecules when the heats of reaction ($\Delta_r H$ or Q) are known:

$$\Delta_r H = \Delta E_0 + \Delta(PV) = \Delta E_0 + \Delta ZPE + \Delta_T H + \Delta nRT \tag{1}$$

where ΔE_0 is the difference in energy between the products and the reactants at 0 K; ΔZPE is the difference between the zero point energies of the products and reactants; and $\Delta_T H$ is the difference between the thermal corrections from 0 to 298.15 K of the products and reactants. $\Delta(PV)$ in the above equation is the PV work term. It equals to ΔnRT for ideal gas reactions. For isodesmic reactions, $\Delta n = 0$.

The optimized structures were employed as the input geometries to determine the densities (ρ) using the CVFF force field and the Ewald summation method in Accelrys's Materials Studio 4.1 software package [23]. Kamlet and Jacob's semi-empirical equations were used to determine the detonation performances of the molecules [24]:

$$D = 1.01\left(NM^{1/2}Q^{1/2}\right)^{1/2}(1 + 1.30\rho) \tag{2}$$

$$P = 1.558\,NM^{1/2}Q^{1/2}\rho^2 \tag{3}$$

where D is the detonation velocity in km/s, P is the detonation pressure in GPa, N is the number moles of gaseous detonation products per gram of explosive, M is the average molecular weight of the gaseous products, Q is the energy of the explosion in kcal/g of explosive, and ρ is the

density in g/cm^3. The possible detonation products were determined based on the modified Kistiakowsky–Wilson rules [25].

Results and discussion

Heat of formation

We performed structure optimizations for the title compounds at the B3LYP/aug-cc-pVDZ level, and their molecular frameworks are shown in Fig. 1. All of the model compounds belong to the C_1 point group. The geometries, bond lengths and total energies were found to be vary with the position of the NH_2 group in the diazole framework. The total molecular energies, zero point energies, thermal corrections to the enthalpy and the heats of formation of the model compounds along with the reference compounds imidazole, methane, nitromethane and aminomethane computed at the B3LYP/aug-cc-pVDZ level are summarized in Table 1. It is known that choosing appropriate reference compounds in the isodesmic reaction leads to good agreement of the calculated $\Delta_f H$ values of the molecules with their experimental values. To the best of our knowledge, there are no experimental heats of formation for the designed compounds reported in the literature. The isodesmic reaction for estimating the heats of formation of the molecules at 298 K can be written as shown in Scheme 1.

The calculated heats of formation of the model compounds were superior ($\approx$180–354 kJ mol^{-1}) to the experimental heats of formation of 2,4-DNI (20.5 kJ mol^{-1}), MTNI (170.3 kJ mol^{-1}), RDX (70.63 kJ mol^{-1}) and HMX (74.88 kJ mol^{-1}), so the model compounds appear to rather promising. The highest and lowest heats of formation among the aminonitroimidazoles were obtained for R66 and R61: 260 and 180 kJ mol^{-1}, respectively. The model trinitroimidazole molecules R30, R32, R34 and R36 had the highest $\Delta_f H$ values (294.3, 338.04, 336 and 353.8 kJ mol^{-1}, respectively). The $\Delta_f H$ values of R30, R32, R65 and R68 obtained at the B3P86/aug-cc-pVTZ level of calculation were 298.5, 340.15, 192.0 and 193.6 kJ mol^{-1}, respectively, thus showing consistency in the values that are obtained at the B3LYP/aug-cc-pVDZ level. The discrepancies in $\Delta_f H$ or Q values among the isomers are caused by the relative positions of the NH_2 and NO_2 groups in the molecules. The heats of formation are highly positive, which is one of the characteristics of energetic materials. Note, however, that the $\Delta_f H$ values were obtained for the gaseous compounds, whereas they should really be evaluated for the solid phase, which would diminish the magnitudes of the $\Delta_f H$ values.
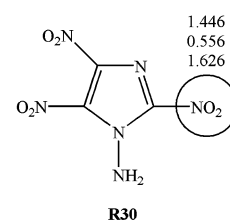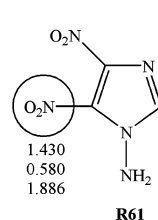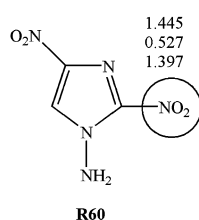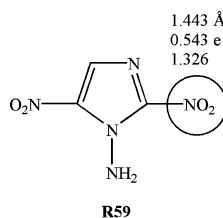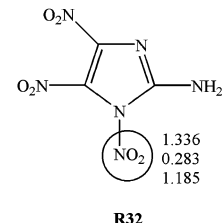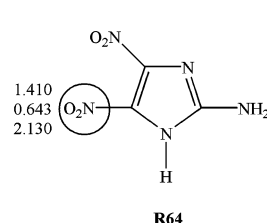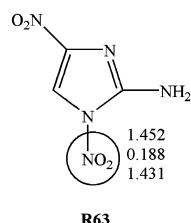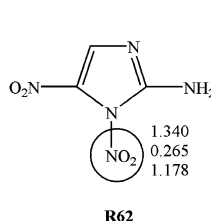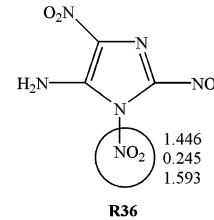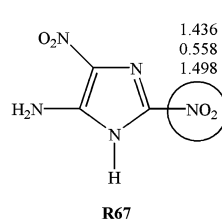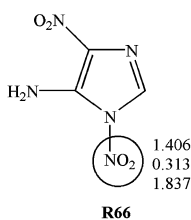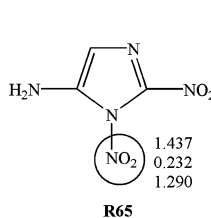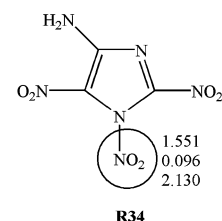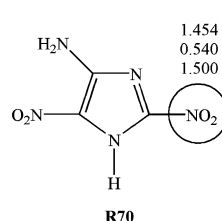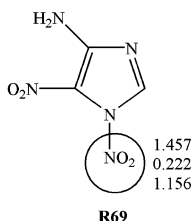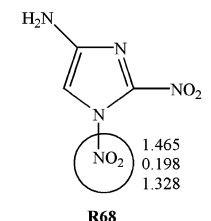
Fig. 1 Molecular structures of the designed compounds, with trigger linkages *encircled*; the *list of values beside each structure* shows the trigger length (in Å), nitro group charge (in e) and midpoint electrostatic potential

Density

The accurate prediction of crystal density is known to be difficult. However, predictive studies of crystal packing are known to give reliable results. The optimized structures computed at the B3LYP/aug-cc-pVDZ level were employed to determine the crystal densities using the Materials Studio 4.1 software package with the CVFF force field and the Ewald summation method. Most of the organic compounds crystallize in the C2/c, P2$_1$, P2$_1$/c, P2$_1$2$_1$2$_1$, P-1, Pbca, Pbcn, Pna2$_1$, CC and C2 space groups. It is known that the density of imidazole increases as the number of NO$_2$ groups increases. Furthermore, higher densities can be achieved if the molecule has an NH$_2$ group that participates in intra- or intermolecular hydrogen bonding with NO$_2$ groups or pyridine-like nitrogen atoms in the same or an adjacent molecule. Table 2 shows the calculated lowest-energy

crystal-density characteristics of aminonitroimidazoles. The model molecules were found to pack in four space groups: P2$_1$/c, P2$_1$2$_1$2$_1$, Pbca and Pna2$_1$. The experimental densities of 2,4-dinitro-1H-imidazole (2,4-DNI) and 1-methyl-2,4,5-trinitroimidazole (MTNI) are 1.76 and 1.78 g cm^{-3} respectively. The compounds R61, R62, R64, R65, R68 and R69 have the highest densities ($\approx$1.78 g cm^{-3}) among the aminonitroimidazoles. The densities of R30, R32, R65 and R68 optimized at the B3P86/aug-cc-pVTZ level of calculation are 1.93, 1.94, 1.80 and 1.78 g/cm^3, respectively, which are similar to the values obtained for the molecules optimized at the B3LYP/aug-cc-pVDZ level. The NH$_2$ group increases the densities from 1.78 to 1.94 g cm^{-3} in the trinitroimidazoles (R30, R32, R34 and R36). The errors in the calculated densities of these compounds are expected to be less than 0.03 g/cm^3, and are thus these values are known to be useful for calculating performance properties [26, 27].

Table 1 The lowest frequencies (ωL), total energies (E_0), zero-point energies (ZPE), thermal corrections to enthalpy (H_T) and gas-phase heats of formation ($\Delta_f H$) of aminonitroimidazoles and some reference compounds, as computed at the B3LYP/aug-cc-pVDZ level

HEM	$\omega L(cm^{-1})$	E_0(a.u.)	ZPE(a.u.)	H_T(a.u.)	$\Delta_f H$(kJ mol^{-1})
R59	52.8922	−690.6342393	0.092624	0.103251	192.90
R60	55.6189	−690.6371644	0.092414	0.103428	199.92
R61	25.7308	−690.6296358	0.092511	0.103636	180.00
R62	75.0277	−690.6490001	0.091668	0.102634	226.02
R63	18.1811	−690.6264625	0.089372	0.098384	185.08
R64	17.8044	−690.6814202	0.092186	0.103552	235.00
R65	57.7604	−690.529637	0.091123	0.102200	189.14
R66	72.1146	−690.659376	0.092025	0.102773	259.96
R67	68.8488	−690.692030	0.092096	0.103311	206.43
R68	36.6195	−690.636329	0.090945	0.102227	191.20
R69	79.3520	−690.6538497	0.091989	0.102847	245.27
R70	65.0897	−690.6958847	0.099245	0.102847	197.00
R30	26.4991	−895.1412989	0.094518	0.107953	294.30
R32	36.0284	−895.636672	0.093638	0.107418	338.04
R34	23.4467	−895.1622831	0.093041	0.107052	336.00
R36	52.8106	−895.1692443	0.093840	0.107239	353.78
Imidazole	549.6127	−226.2500375	0.070976	0.075666	132.5[a]
CH$_4$	1303.53	−40.520622	0.044261	0.048080	−74.6[a]
CH$_3$NO$_2$	65.8541	−245.054716	0.049551	0.054835	−80.8[a]
CH$_3$NH$_2$	307.812	−95.873895	0.063532	0.067919	−22.5[a]

[a] Experimental values from [22]

Detonation performance

The detonation velocity (D) and pressure (P) are the important parameters for evaluating the performance and applicability of each compound. As the number of NO$_2$ groups on the imidazole, the density, the detonation velocity and the detonation pressure increase, so do the oxygen balance and heat of formation. When the number of NO$_2$ groups in the molecule is three, D and P reaches its maximum value for the nitroimidazoles. Also, when there are four NO$_2$ groups in each molecule, Q and D decrease, because the oxygen balance is too positive, and deviates to far from the perfect oxygen balance. Therefore, it is clear that Q, D and P are maximum when the number of NO$_2$ groups on the imidazole is three. The calculated detonation characteristics of the designed compounds are summarized in Table 3. The calculated detonation pressures of the aminodinitroimidazoles vary from 26.0 to 28.0 GPa, but their detonation velocities are nearly equal. The experimen-

tal detonation velocity and pressure of 2,4-DNI are 8.13 km/s and 28.10 GPa, respectively. The detonation velocity and pressure of MTNI are 8.96 km/s and 32.54 GPa, respectively. The calculated performance properties of 2,4-DNI and MTNI agree with their experimental values, so the predictions for the title compounds are observed to be reliable. The compounds R64, R67 and R70 may find similar application to 4-amino-3,5-dinitro-1H-pyrazole (LLM-116) due to their promising properties as potential insensitive munitions. The calculated densities of the isomeric compounds are close to each other, with deviations of 0.02–0.06 g/cm^3, so D and P are also similar for these compounds, as they are mainly determined by ρ and ρ^2, respectively [24]. The detonation characteristics of aminotrinitroimidazoles were found to be superior than those of RDX (D 8.75 km/s, P 34.70 GPa) and HMX (D 8.96 km/s, P 35.96 GPa). The performance properties of the model molecules R30 (D 9.20 km/s, P 39.42 GPa), R32 (D 9.16 km/s, P 38.66 GPa), R65 (D 7.87 km/s, P 27.53 GPa)
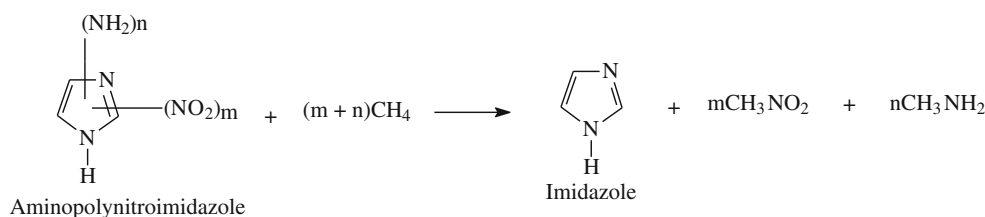
Scheme 1 Isodesmic reaction for estimating the heats of formation of the molecules at 298 K

Table 2 Crystal density characteristics of aminonitroimidazoles computed at the B3LYP/aug-cc-pVDZ level

HEM	Cell volume	Total energy (kJ mol^{-1})	Space group	Crystal system	a, b, c (Å) α, β, γ	Density (g cm^{-3})
R59	658.54892568	45.63597178	P2$_1$/c	Monoclinic	6.34, 9.47, 13.38 $\alpha = \gamma = 90°$, $\beta = 125°22'$	1.75
R60	656.01869175	35.22331902	Pna2$_1$	Orthorhombic	9.32, 6.51, 10.817 $\alpha = \beta = \gamma = 90°$	1.75
R61	643.21613440	−4.63548260	P2$_1$/c	Monoclinic	17.28, 9.16, 7.97 $\alpha = \gamma = 90°$, $\beta = 56°36'$	1.78
R62	646.66767983	−65.26061201	Pna2$_1$	Orthorhombic	8.63, 7.67, 9.76 $\alpha = \beta = \gamma = 90°$	1.77
R63	662.72194457	−52.44768440	Pna2$_1$	Orthorhombic	13.58, 6.42, 9.76 $\alpha = \beta = \gamma = 90°$	1.74
R64	644.81372053	−46.03492401	P2$_1$2$_1$2$_1$	Orthorhombic	6.67, 6.04, 17.60 $\alpha = \beta = \gamma = 90°$	1.78
R65	646.71190843	−26.12259996	P2$_1$/c	Monoclinic	7.63, 8.30, 12.97 $\alpha = \gamma = 90°$, $\beta = 127°98'$	1.78
R66	658.77981276	−29.3842280	P2$_1$/c	Monoclinic	18.07, 9.56, 13.64 $\alpha = \gamma = 90°$, $\beta = 163°77'$	1.74
R67	658.58439721	18.059133246	P2$_1$2$_1$2$_1$	Orthorhombic	10.57, 6.31, 9.88 $\alpha = \beta = \gamma = 90°$	1.75
R68	647.01323524	−27.74764905	P2$_1$2$_1$2$_1$	Orthorhombic	10.63, 5.20, 11.71 $\alpha = \beta = \gamma = 90°$	1.78
R69	657.752311	−44.1286101	Pna2$_1$	Orthorhombic	8.37, 7.62, 10.10 $\alpha = \beta = \gamma = 90°$	1.78
R70	657.7523153	27.33391136	P2$_1$/c	Monoclinic	10.36, 6.46, 9.94 $\alpha = \gamma = 90°$, $\beta = 81°81'$	1.75
R30	754.98698573	24.5616875	P2$_1$/c	Monoclinic	6.70, 16.41, 7.14 $\alpha = \gamma = 0°$, $\beta = 99°84'$	1.94
R32	1498.3101	−106.73923675	Pbca	Orthorhombic	16.1, 9.87, 9.42 $\alpha = \beta = \gamma = 90°$	1.93
R34	1498.8977	−58.70938566	Pbca	Orthorhombic	16.03, 9.85, 9.48 $\alpha = \beta = \gamma = 90°$	1.93
R36	748.05680	−44.63809745	P2$_1$/c	Monoclinic	17.11, 6.75, 6.50 $\alpha = \gamma = 90°$, $\beta = 59°45'$	1.94

and R68 (D 7.94 km/s, P 27.55 GPa) obtained at the B3P86/aug-cc-pVTZ level of calculation appear to be similar to those computed at the B3LYP/aug-cc-pVDZ level.

Impact sensitivity correlations

We also approximated the impact sensitivities ($h_{50\%}$) of the model compounds from the electronic structures using a Mulliken atomic charge analysis of the nitro (NO$_2$) groups. In nitro compounds, the C–NO$_2$, N–NO$_2$ and O–NO$_2$ bonds are usually the weakest bonds in the molecule, and breaking one of these bonds is often the initial step in the decomposition or detonation. The nitro group charge ($-Q_{NO2}$) can be calculated by summing the net Mulliken atomic charges on the nitrogen (Q_N) and oxygen (Q_{O1} and Q_{O2}) atoms:

$$-Q_{NO2} = Q_N + Q_{O1} + Q_{O2}. \tag{4}$$

The higher the negative charge on the NO$_2$ group, the weaker the electron-withdrawing ability and thus the greater the overall stability of the compound. As the number of NO$_2$ groups on the imidazole framework increases, competition for the available charge increases and thus the molecule becomes unstable. In other words, when there is one nitro group on imidazole (a mononitroimidazole), it will have a larger negative charge than if there are two or more nitro groups (a polynitroimidazole). Therefore, more sensitive compounds will have smaller charges on their nitro groups. The calculated V_{mid} values were found to vary from 1.326 to 2.12. The computed $-Q_{NO2}$ values, midpoint electrostatic potentials (V_{mid}), and trigger lengths of the molecules are presented in Table 4. The calculated $-Q_{NO2}$ values of the model compounds, aside from R63 (0.188e), R68 (0.198e) and R34 (0.096e), vary from 0.232e to 0.651e. As approximated by Zhang et al. [15, 16], these values are higher than those of RDX (0.134e), HMX (0.112e), TNAZ

Table 3 The densities (ρ), chemical energies of detonation (Q), detonation velocities (D), and detonation pressures (P) of the model compounds along with those of the known explosives 2,4-DNI, MTNI, RDX and HMX

HEM	ρ(g cm^{-3})	Q(kcal g^{-1})	D(km s^{-1})	P(GPa)
R59	1.75	0.86	7.83	26.70
R60	1.75	0.87	7.84	26.83
R61	1.78	0.84	7.91	27.63
R62	1.77	0.98	8.02	28.33
R63	1.74	0.85	7.77	26.22
R64	1.78	0.94	8.05	28.53
R65	1.78	0.86	7.92	27.52
R66	1.74	0.95	8.00	27.76
R67	1.75	0.90	7.84	26.67
R68	1.78	0.86	7.92	27.57
R69	1.78	0.93	8.00	27.80
R70	1.75	0.87	7.84	26.77
R30	1.93	1.42	9.23	39.41
R32	1.93	1.37	9.14	38.65
R34	1.93	1.37	9.15	38.70
R36	1.94	1.34	9.12	38.53
2,4-DNI	1.80 (1.76)[b]	0.87	8.10 (8.13)	29.04 (28.10)
MTNI	1.86 (1.78)	1.07	8.66 (8.96)	33.92 (32.54)
RDX	1.78 (1.82)	1.25 (1.27)	8.86 (8.75)	34.23 (34.70)
HMX	1.90 (1.92)	1.25 (1.27)	9.10 (8.96)	39.40 (35.96)

[b] Experimental values from [6]

(0.114e), CL-20 (0.081e), and ONC (0.146e), and the compounds are found to be more insensitive. This impact sensitivity or stability is attributed to the presence of a π-excessive aromatic heterocyclic ring, the delocalization of π-electrons, and the presence of new types of intramolecular N–H…O and N…O interactions.

Electric spark sensitivity correlations

Electric spark sensitivity is the degree of sensitivity of an energetic compound to an electrical discharge. It is usually determined by subjecting the explosive compound to a high-voltage discharge from a capacitor. Zeman [17, 18] gave a relationship between the electric spark sensitivity (E_{ES}), the square of the detonation velocity (D), the reciprocal temperature, the Piloyan activation energy, and the heat of fusion of an energetic compound. Wang et al. [28, 29] provided a suitable correlation between the electric spark sensitivity, the detonation velocity and the detonation pressure for nitramines and nitroarenes. We have approximated the electric spark sensitivities of model compounds based on the lowest unoccupied molecular orbital energies and nitro group charge analysis, as per the procedure in [19]:

$$E_{ES}(J) = (-1)^{n1} 10.16 Q_{NO2} - 1.05_{n1\,n2} E_{LUMO} - 0.20, \quad (5)$$

where, $n1$ is the number of aromatic rings, $n2$ is the number of substituents attached to the imidazole, such as alkyl (−R)

or amino (−NH$_2$) groups, −Q_{NO2} is the minimum Milliken charge on the NO$_2$ group and E_{LUMO} (in eV) is the energy of the lowest unoccupied molecular orbital (LUMO). As calculated by Zhi et al. [19], the calculated electric spark sensitivity values of the designed compounds R32 (7.12 J), R36 (7.08 J) and R66 (7.31 J) are similar to those of TNT (7.0 J), while the compounds R62, R63, R65, R68, R69 and R34 show lower electric spark sensitivities than TNT. On the other hand, the model compounds R61, R64 and R30 possess exceptionally high electric spark sensitivities of 10.76, 10.16 and 10.50 J, respectively.

Frontier molecular orbital energies

The frontier molecular orbital energies and the gaps between them are presented in Table 4. Fukui et al. [30] was the first to notice the prominent role played by the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) in governing the reactivity of aromatic compounds. The smaller the band gap between the HOMO and the LUMO of the compound, the easier it is for electrons to cross this gap, and so the poorer the stability of the compound. The LUMO energies vary from −0.11993 to −0.148322 a.u. and the HOMO energies range from −0.31312 to −0.27720 a.u. for aminodinitroimidazoles. The LUMO energies vary from −0.16401 to −0.14180 a. u. and the HOMO energies from −0.30820 to −0.24902 a.u. for aminotrinitroimidazoles. The frontier orbital energy gaps vary from 0.16962 to 0.10658 a.u. for aminodinitroimi-

Table 4 Trigger lengths, nitro group charges ($-Q_{NO2}$), midpoint electrostatic potentials (V_{mid}), electric spark sensitivities (E_{ES}), frontier orbital energies, and the gaps between these orbitals for the model compounds

HEM	Bond	Length(Å)	$-Q_{NO2}$(e)	V_{mid}	E_{ES}(J)	LUMO(a.u.)	HOMO(a.u.)	$\Delta E_{(LUMO-HOMO)}$(a.u.)
R59	C2–NO$_2$	1.44298	0.543	1.32642	9.95467	−0.14832	−0.29525	0.14693
	C5–NO$_2$	1.42594	0.586	1.76305	10.39155			
R60	C2–NO$_2$	1.44556	0.527	1.39738	9.51354	−0.13857	−0.30819	0.16962
	C4–NO$_2$	1.45259	0.571	1.64121	9.96057			
R61	C4–NO$_2$	1.46473	0.651	1.55878	10.76766	−0.13837	−0.30103	0.16266
	C5–NO$_2$	1.42858	0.586	1.88578	10.10726			
R62	N1–NO$_2$	1.33953	0.265	1.17802	6.47532	−0.12540	−0.26801	0.14261
	C5–NO$_2$	1.42390	0.646	2.09003	10.34628			
R63	N1–NO$_2$	1.45213	0.188	1.43100	6.17988	−0.14244	−0.24902	0.10658
	C4–NO$_2$	1.46933	0.654	1.69601	10.91434			
R64	C4–NO$_2$	1.46658	0.687	1.80965	10.60656	−0.11993	−0.26771	0.14778
	C5–NO$_2$	1.41041	0.643	2.12705	10.15951			
R65	N1–NO$_2$	1.43721	0.232	1.28861	6.34095	−0.13243	−0.25826	0.12583
	C2–NO$_2$	1.44161	0.560	1.34295	9.07343			
R66	N1–NO$_2$	1.40640	0.313	1.83731	7.31099	−0.13758	−0.26158	0.1240
	C4–NO$_2$	1.41672	0.712	1.58958	11.36483			
R67	C2–NO$_2$	1.43605	0.558	1.49855	9.56652	−0.12940	−0.26820	0.1388
	C4–NO$_2$	1.42317	0.753	1.02096	11.54769			
R68	N1–NO$_2$	1.46530	0.198	1.32880	6.05705	−0.13423	−0.25296	0.11873
	C2–NO$_2$	1.47120	0.535	1.34375	9.47081			
R69	N1–NO$_2$	1.45701	0.222	1.15580	5.92126	−0.12201	−0.26803	0.14602
	C5–NO$_2$	1.40348	0.740	2.35701	11.20446			
R70	C2–NO$_2$	1.45379	0.540	1.49952	9.81676	−0.14456	−0.26636	0.12180
	C5–NO$_2$	1.39678	0.780	2.56017	12.25517			
R30	C2–NO$_2$	1.44628	0.551	1.62624	10.49764	−0.16401	−0.31312	0.14911
	C4–NO$_2$	1.46540	0.642	1.78792	11.41998			
	C5–NO$_2$	1.43730	0.640	1.87574	11.39965			
R32	N1–NO$_2$	1.33627	0.283	1.18538	7.12651	−0.14179	−0.28739	0.14560
	C4–NO$_2$	1.46933	0.650	1.83893	10.85522			
	C5–NO$_2$	1.42609	0.615	1.98164	10.49962			
R34	N1–NO$_2$	1.55116	0.096	2.12986	5.57601	−0.15402	−0.27719	0.12317
	C2–NO$_2$	1.45791	0.535	1.55153	10.03625			
	C4–NO$_2$	1.40034	0.717	2.28516	11.88537			
R36	N1NO$_2$	1.44581	0.245	1.15934	7.08528	−0.15386	−0.28583	0.13197
	C2–NO$_2$	1.45277	0.537	1.48957	10.05201			
	C5–NO$_2$	1.42433	0.703	1.88441	11.73856			

dazoles, and from 0.12317 to 0.14911 a.u. for amino-trinitroimidazoles, which indicates that the model compounds are stable. For the aminodinitroimidazoles, the band gap of R60 is the largest (0.16962 a.u.) and R63 is the smallest (0.10658 a.u.), indicating that the former is more stable than the latter. The compounds R30, R32, R59, R60, R62, R63 and R64 appear to the most stable, where "stability" here refers to chemical or photochemical processes involving electron transfers or electron jumps. Therefore, the stability of the aminodinitroimidazoles decreases as follows: R60>R61>R2~R64>R59>R69>R62>R65>R66>R70>R68>R63. The corresponding order for the aminotrinitroimidazoles is: R30>R32>R36>R34. C–NO$_2$ and N–NO$_2$ are the trigger bonds, and the

resonance of the imidazole ring strengthens these bonds, thereby stabilizing the molecules. The longer the trigger bonds of the molecule, the easier it is to break them, so the molecule becomes less stable. Based on the total energies of the compounds concerned, the most stable and the least stable compounds among the aminodinitroimidazoles are R65 (−690.529637 a.u.) and R70 (−690.6958847 a.u.), respectively. The most stable and the least stable compounds among the aminotrinitroimidazoles are R30 (−895.141299 a.u.) and R36 (−895.1692443 a.u.), respectively. The higher the total energy of the molecule, the less stable the. The discrepancies in the stabilities of the molecules are due to slight variations in total energies, trigger bond lengths, and band gap values among the

isomers, which in turn are caused by changes in the relative positions of the nitro and amino groups.

Conclusions

In conclusion, the model compounds satisfied the criteria for high energy density compounds. The energies of detonation (Q), densities (ρ), detonation velocities (D) detonation pressures (P), stabilities, impact insensitivities, and electric spark insensitivities increased markedly as the number of hydrogens substituted on the nitroimidazoles increased. The density ($\rho \approx 1.93$ g/cm^3) and detonation properties ($D \approx 9.10$ km/s and $P \approx 38.50$ Gpa) of the aminotrinitroimidazoles appear to be promising compared with those of RDX and HMX. The discrepancies in the performance properties and stabilities or sensitivities of the compounds are caused by the relative positions of the NH_2 and NO_2 groups. The calculated heats of formation, densities, detonation velocities, detonation pressures, and impact and electric spark sensitivities of the model compounds are crude approximations, so there may be errors in these calculated values when compared to the real ones.

References

1. Larina L, Lopyrev V (2009) Nitroazoles: synthesis, structure and applications. Springer, Berlin
2. Bulusu S, Damavarapu R, Autera JR, Behrens R, Minier LM Jr, Villanueva J, Jayasuriya K, Axenord T (1995) Thermal reaargenement of 1,4-dinitroimidazole to 2,4-dinitroimidazole: characterization and investigation of the mechanism by mass spectrometry and isotope labeling. J Phys Chem 99:5009–50015
3. Damavarapu R, Jayasuriaya K, Vladimroff T, Iyer S (1995) 2,4-Dinitroimidazole: a less sensitive explosive and propellant made by thermal rearrangement of molten 1,4-dinitroimidazole. US Patent 5,387,297
4. Novikov SS, Khmel'nistkii LI, Lebedev OV, Sevast'yanova VV, Epishina LV (1970) Nitration of imidazoles with various nitrating agents. Chem Heterocycl Compd 6:465–469
5. Katritzky AR, Cundy DJ, Chen J (1993) Polyiodoimidazoles and their nitration products. J Energetic Mater 11:345–352
6. Damavarapu R, Surapaneni RC, Gelbel N, Duddu R, Zhang M, Dave PR (2007) Melt-cast explosive material. US Patent 7,304,164
7. Jadhav HS, Talawar B, Sivabalan R, Dhavale DD, Asthana SN, Krishnamurthy VN (2007) Synthesis, caracterization and thermolysis studies on new derivatives of 2,4,5-trinitrimidazoles: potential insensitive high energy materials. J Hazad Mater 143:192–197
8. Kamlet MJ, Adolph HG (1979) The relationship of impact sensitivity with structure of organic high explosives. II. Polynitroaromatic explosives. Propellants Explos Pyrotech 4:30–34
9. Politzer P, Murray JS (1995) C-NO$_2$ dissociation energies and surface electrostatic maxima in relation to impact sensitivities of some nitroheterocyclic molecules. Mol Phys 86:251–255
10. Murray JS, Concha MC, Politzer P (2009) Links between the surface electrostatic potential of energetic molecules, impact sensitivities and C–NO$_2$/N–NO$_2$ bond dissociation energies. Mol Phys 107:89–97
11. Rice BM, Hare JJ (2002) A quantum mechanical investigation of the relation between the impact sensitivity and charge distribution in energetic molecules. J Phys Chem A 106:770–1783
12. Murray JS, Lane P, Politzer P (1995) Relationship between impact sensitivities and molecular surface electrostatic potentials in nitroaromatic and nitroheterocyclic compounds. Mol Phys 85:1–8
13. Murray JS, Lane P, Politzer P (1998) Effects of strongly electron attracting components on the molecular electrostatic potentials: application to predicting the impact sensitivities of energetic molecules. Mol Phys 93:187–194
14. Pospìšil M, Vávra P, Concha MC, Murray JS, Politzer P (2010) A possible crystal volume factor in the impact sensitivities of some energetic compounds. J Mol Model 16:895–901
15. Zhang C, Shu Y, Huang Y, Zhao X, Dong H (2005) Investigation of correlation between impact sensitivites and nitro group charges in nitro compounds. J Phys Chem B 109:8978–8982
16. Zhang C (2009) Review of the establishment of the nitro group charge method and its applications. J Hazard Mater 161:21–28
17. Zeman S (1999) Relation between the detonation characteristics and ^{15}N NMR chemical shifts of nitramines. J Energetic Mater 17:305–330
18. Zeman S (2006) New aspects of initiation reactivities of energetic materials demonstrated in nitramines. J Hazard Mater 132:155–164
19. Zhi C, Cheng X (2010) The correlation between electric spark sensitivity of polynitroaromatic compounds and their molecualr electronic properties. Propellants Explos Pyrotech. doi:10.1002/prep.20090092
20. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery AJ, 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 J, 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 B.04. Gaussian Inc., Pittsburgh
21. Hahre WJ, Radom L, Schleyer PvR, Pole JA (1986) Ab initio molecular orbital theory. Wiley, New York
22. David RL (2003–2004) CRC handbook of chemistry and physics, 84th edn. CRC Press, Boca Raton
23. Accelrys Inc. (2004) Materials Studio 4.01. Accelrys Inc., San Diego
24. Kamlet MJ, Jacobs SJ (1968) Chemistry of detonation. I. A simple method for calcualting the detonation proiperties of C–H–N–O explosives. J Chem Phys 48:23–35
25. Akhavan J (1998) Chemistry of explosives. Royal Society of Chemistry, Cambridge
26. Kim JK, Cho SG, Kim CK, Park H-Y, Zhang H, Lee HW (2008) Prediction of densities for solid energetic molecules with molecular surface potentials. J Comput Chem J Comput Chem 29:1818–1824

27. Politzer P, Martinez J, Murray JS, Concha MC, Alejandro TB (2009) An electrostatic interaction correction for improved crystal density prediction. Mol Phys 107:2095–2101

28. Wang G, Xiao H, Ju X, Gong X (2006) Calculation of detonation velocities and pressures and their relationship with the electric spark sensitivities of nitramines. Propellants Explos Pyrotech 31:102–109

29. Wang G, Xiao H, Ju X, Gong X (2006) Calculation of detonation velocity, pressure and electric spark sensitivity of nitro arenes based on quantum chemistry. Propellants Explos Pyrotech 31:361–368

30. Fukui F, Yonezawa T, Shingu HJ (1952) A molecular orbital theory of reactivity in aromatic hydrocarbons. J Chem Phys 20:722–725

ORIGINAL PAPER

An information-carrying and knowledge-producing molecular machine. A Monte-Carlo simulation

Christoph Kuhn

Abstract The concept called *Knowledge* is a measure of the quality of genetically transferred information. Its usefulness is demonstrated quantitatively in a Monte-Carlo simulation on critical steps in a origin of life model. The model describes the origin of a bio-like genetic apparatus by a long sequence of physical-chemical steps: it starts with the presence of a self-replicating oligomer and a specifically structured environment in time and space that allow for the formation of aggregates such as assembler-hairpins-devices and, at a later stage, an assembler-hairpins-enzyme device—a first translation machine.

Keywords Assembler-hairpins-enzyme device ·
Emergence and storage of information · Genetic apparatus ·
Knowledge · Monte-Carlo · Origin of life · Structured
time-space environment

Introduction

Modeling the origin of life: postulates on initial conditions

This paper is based on the concept that living individuals—distinct aggregates of interlocking molecules—are a form of matter that carries information to be reproduced and to evolve in their given environment into forms of increasing complexity and intricacy [1–7].

C. Kuhn (✉)
Biomedical Optics Research Laboratory, Clinic of Neonatology,
University Hospital Zürich,
Frauenklinikstrasse 10,
8091 Zurich, Switzerland
e-mail: c-k@gmx.ch

A step-by-step Darwinian process appears instantaneous when the first entity emerges by chance (a de-novo oligomer that is capable of replication), survives as a species (population of individuals of a certain kind) and then evolves by continued reproduction with variation and selection in the particular environment driving the process. The initial step, which has been called "origin of life", requires a basic initial form for the building blocks as well as basic initial conditions for the environment (Supplementary Fig. 1). Such a step-by-step Darwinian process, continuous and sustained, leads first to aggregates of interlocking molecules, and then to the bio-genetic apparatus.

These aggregates reach increasing independence from the afore-mentioned highly specific initial conditions by gradually populating a diversified area, as follows: (1) populating regions of compartments with increasing size of pores, (2) evolving devices that produce envelopes as innate compartments, (3) evolving metabolism (increasing intricacy of the living machinery).

Modeling origin of life: the RNA world

We consider the case where only one type of complementary R-monomers, R_1 and R_2, is present, and we assume that de-novo oligomers emerge and are capable of replicating (Supplementary Fig. 2a). The initial sequence is random and, due to errors in the copying process, there is variation of this sequence. If by chance a strand is found in the population that has a sequence such that it folds onto itself to form a hairpin (Supplementary Fig. 2c), the hairpin will be selected because it is more protected against hydrolysis and thus has a higher probability of being selected. The hairpin finds an open strand to bind by the loop by complementary and anti-parallel means. If the number of adjacent hairpins equals or exceeds three, such an aggregate will be selected because it is, again, more protected against hydrolysis, and thus has a higher probability of being selected. A reading

frame (Supplementary Fig. 2d) exists if all hairpins in the population have loops of one kind. All hairpins bound to the assembler are then adjacent (Supplementary Fig. 2e).

The early process (RNA-world [8–11]) comes to an end when replication reaches a certain precision. This apparent dead-end is overcome by the emergence of the first translation apparatus with an enzyme E_1 as its product (i.e., the assembler-hairpins-enzyme device, the HAE_1-device and its nonsensical replica, the HAE_0-device as explained below).

Modeling the origin of life: the RNA–protein world

We introduce in our model two a-monomers, a_1 and a_2, (which may be glycine and alanine). Occasionally, by errors in the copying process, the two R-monomers at the open end (Supplementary Fig. 3a) of the hairpin are no longer complementary. The hairpin then carries an a-monomer such that R_1–R_1 carries a_1 and R_2–R_2 carries a_2 (Supplementary Fig. 3a). If all hairpins on an assembler carry a-monomers, these can then oligomerize. The product, an a-oligomer with random sequence, has no enzymatic power but functions as an agglutinate-forming envelope E_0 (Supplementary Fig. 3b).

The HAE_1-device is a translation apparatus (Supplementary Fig. 3c), and emerges as a by-product. It is selected only if (1) there is a code on the (+)-assembler for this specific a-oligomer, (2) this code is translated by the hairpins, and (3) the specific a-oligomer is an enzyme (enzyme E_1) that increases the precision of the R-replication (replicase). The (−)-assembler constitutes the anti-parallel complimentary copy of an (+)-assembler, and its product therefore has no sense (HAE_0-device).

Increasing contamination coupled with increasing complexity of the evolving system (i.e., by accumulation of HAE_0-devices with an increasing number of different HAE_i-devices) again leads to another barrier. This barrier is overcome by a fundamental change in the machinery into a primordial form of the translation apparatus (DNA–RNA–protein world [12–16]). The emergence of a bio-genetic apparatus [7] then paves the way for the "explosion of life" [17–19].

Computer implementation and simulation

Supplementary Fig. 4 presents the computer implementation [20, 21] in a flow chart overview format. The top part shows the construction phase, where the aggregates are formed. According to the fitness of the formed devices, parameters such as survival chance, replication probability and the probability of occurrence of an error in the replication, are assigned to these entities. Possible aggregates are the hairpin (Supplementary Fig. 2c), the HA-device (Supplementary Fig. 2e) and, after emergence of new kinds of monomers (a-monomers) that attach to the open ends of hairpins, the

HAE_0- (Supplementary Fig. 3a) and the HAE_1- (Supplementary Fig. 3c) devices. In the selection phase (middle of Supplementary Fig. 4), only a fraction of the population survives (corresponding to half of the total number of strands). In the multiplication phase (bottom of Supplementary Fig. 4), the aggregates dissociate and the strands are copied by chance until the total number of strands is replenished. Two typical simulations starting with a single strand of random sequence are shown in Supplementary Fig. 5a,b.

Quality of information *Knowledge* demonstrated by a Monte Carlo study

Defining quantities that measure *Information* and *Knowledge* elucidates some principle aspects of the origin of life. Genetic information is stored in one entity and transferred from one generation to the next as a distinct sequence of monomers in a strand. It is measured in bits [22].

"The quality of genetic information" denotes that the genetic information has the property to instruct the formation of an entity that behaves as if it had knowledge, that is, it behaves as if it would know how to survive and to multiply in its given environment.

Knowledge, K, is related to the effort required to develop a certain degree of functionality. Let us consider the total information I(g) (number of bits at generation g) that has to be discarded by eliminating unfit individuals along a singlet out trial of a Darwinian evolution that takes place within a population of entities from the initial replicating oligomer at generation 0, continued by many step-by-step Darwinian processes, until the given degree of functionality is reached at generation g. The average g_{bt} of the repeated process (Monte Carlo method) levels out random fluctuation in g. The effort required to develop the given degree of functionality is measured by the number $K = I(g_{bt})$ called *Knowledge* K [3, 7, 23].

Beginning with a situation in which the emergence of a first replicating strand is possible, it takes a huge number of trials (i.e., discarding much information) until a replicable strand actually appears (*Knowledge* K growing suddenly from zero to a value K_0 given by the sum of discarded bits (see schematic illustration in Supplementary Fig. 6a, b). Then, *Knowledge* K will stepwise increase, being constant (reflecting the gradual adaptation of the evolving forms to a given region by refinements of their organizational structure) until a major change in the organizational structure appears with a stepwise increase in *Knowledge* K (breakthrough process leading to the colonization of a new region). Supplementary Fig. 7a (analyzing the emergence of hairpins) and Supplementary Fig. 7b (showing *Knowledge* K of the evolution from replicating strands to the

HAE$_1$-device) are evaluated from the statistics of the Monte Carlo study. Building up the HA-device with its reading frame requires many generations, whereas the HAE$_1$-device emerges shortly after a-monomers appear that are able to link to the hairpin's open end.

Conclusions

Critical steps in a model on the origin of a bio-like genetic apparatus have been demonstrated by a computer simulation and analyzed by Monte-Carlo method. Establishing a reading frame, from a state of randomly formed aggregates of hairpins to a state of a well-ordered HA-device required many generations of a step-by-step Darwinian process. Modeling the origin of a bio-like genetic apparatus—from the emergence, de-novo, of a self-replicating oligomer and a specifically structured environment in time and space that allows the formation of aggregates such as assembler-hairpins-devices and continuing evolution to increasing complexity, which ultimately leads to the assembler-hairpins-enzyme device—is important to understand the principles of the origin of life and to stimulate testing the origin of life model by experimental efforts. The concept of *Knowledge* is therefore a useful measure for the quality of information along Darwinian evolution.

References

1. Kuhn H (1972) Self-organization of molecular systems and evolution of the genetic apparatus. Angew Chem Int Ed Engl 11:798–820
2. Kuhn H (1976) Model consideration for the origin of life. Environmental structure as stimulus for the evolution of chemical systems. Naturwissenschaften 63:68–80
3. Kuhn H, Waser J (1983) Self-organization of matter and early evolution of life. In: Hoppe W, Lohmann W, Markl H, Ziegler H (eds) Biophysics. Springer, Berlin, pp 830–874
4. Kuhn H, Waser J (1994) A model of the origin of life and perspectives in supramolecular engineering. In: Behr JP (ed) Lock-and-key principle. Wiley, Chichester, pp 247–306
5. Kuhn H, Kuhn C (2003) Diversified world: Drive of life's origin?! Angew Chem Int Ed 42:262–266
6. Kuhn H (2008) Origin of life—symmetry breaking in the universe: emergence of homochirality. Curr Opin Colloid Interface Sci 13:3–11
7. Kuhn H (2010) Is the transition from chemistry to biology a mystery? J Syst Chem 1:3
8. Orgel LE (1973) The origins of life: molecules and natural selection. Wiley, NewYork
9. Orgel LE (2004) Prebiotic chemistry and the origin of the RNA world. Crit Rev Biochem Mol Biol 39:99–123
10. Dworkin JP, Lazcano A, Miller SL (2003) The roads to and from the RNA-world. J Theor Biol 222:127–134
11. Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature 459:239–242
12. Bartel DP, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences. Science 261:1411–1418
13. Keefe AD, Szostak JW (2001) Functional proteins from a random sequence library. Nature 410:715–718
14. Seelig B, Szostak JW (2007) Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. Nature 448:828–831
15. Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. Nature 409:387–390
16. Shevack A, Gewitz HS, Hennemann B, Yonath A, Wittmann HG (1985) Characterization and crystallization of ribosomal particles from *Halobacterium marismortui*. FEBS Lett 184:68–71
17. Crick FHC (1958) On protein synthesis. Symp Soc Exp Biol XII:139–163
18. Crick FHC (1970) Central dogma of molecular biology. Nature 227:561–563
19. Dawkins R (2004) Ancestor's tale. A pilgrimage to the dawn of life. Weidenfeld & Nicolson, London
20. Kuhn C (2001) Computer-modelling origin of a simple genetic apparatus. Proc Natl Acad Sci USA 98:8620–8626
21. Kuhn C (2005) A computer-glimpse of the origin of life. J Biol Phys 31 C:571–585
22. Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana, Illinois
23. Kuhn H (1988) Origin of life and physics: diversified microstructure inducement to form information-carrying and knowledge-accumulating systems. IBM J Res Dev 32:37–46

Ab initio study of electron transport in 4-(3-nitro-4-tetrafluorophenylthiolate-ethynyl, phenylethynyl) benzenethiolate

Lilia Serrato-Villegas · Marco Gallo ·
Marcos Delgado-Ríos · Maria Teresa Romero ·
Daniel Glossman-Mitnik

Abstract The electron transport of the 4-(3-nitro-4-tetrafluorophenylthiolate-ethynyl, phenylethynyl) benzenethiolate (S-FNPPB-o) molecule assembled in two Au (111) electrodes, was studied using two approaches: in the first approximate approach an electric field was applied to the pure molecule attached to two thiolate ends fixed, and in the second approach we used the nonequilibrium Green's function formalism (NEGF) coupled to DFT to calculate the I-V curve and the voltage dependence of the transmission function in the extended system, molecule plus electrodes. By applying an electric field to the pure molecule plus thiolate ends fixed, and visualizing the changes in the spatial distribution of the frontier molecular orbitals, we can expect based on the continuity of the conduction pathway in electron transport, that if electron transport occurs through the frontier orbitals, only the LUMO orbital would create an open channel for electron transport due to its delocalized nature and large orbital density at the thiolate groups. The NEGF calculations indicate that at applied voltages lower than ±0.8 V, the current is related to transmission values through the tails of the broad LUMO orbital, and since this orbital is the one closer to the Fermi energy, and we observed very low current values in this region, higher current values at positive bias than at negative bias. As the voltage exceeds ±0.8 V the current increases from the contribution of more states from the broadened part of the transmission function from the LUMO orbital, and when the voltage approaches ±2 V, the LUMO+1 orbital enters into the bias window and the current increases again.

L. Serrato-Villegas
Facultad de Ciencias Químicas,
Universidad Autónoma de Coahuila,
Blvd. Venustiano Carranza s/n. Saltillo,
Coahuila, Mexico, C.P. 25200

M. Gallo (✉)
Facultad de Ciencias Químicas,
Universidad Autónoma de San Luis Potosí,
Av. Manuel Nava No. 6 Zona Universitaria,
San Luis Potosí S.L.P., C.P. 78210, Mexico
e-mail: marco.gallo@uaslp.mx

M. Delgado-Ríos
Universidad Autónoma de Ciudad Juárez, Henri Dunant 4016,
Zona Pronaf, Cd. Juárez,
Chihuahua, Mexico, C.P. 32310

M. T. Romero
Facultad de Ciencias Físico-Matemáticas,
Universidad Autónoma de Coahuila,
Conjunto Universitario Camporredondo Edificio "D",
C.P., 25000 Saltillo Coahuila, Mexico

D. Glossman-Mitnik
Centro de Investigación en Materiales Avanzados,
S.C., Miguel de Cervantes 120, Complejo Industrial Chihuahua,
Chihuahua, Mexico 31109

Introduction

The semiconductor industry today is facing a rapid path to the development of small circuits, portable cell phones, small supercomputers, faster internet communications, etc. Without computers, scientists today could not develop new pharmaceuticals, track climate changes, develop nanosensors, or image the human brain. In the 1970s Gordon Moore, one of the founders of Intel, predicted that the number of components crammed into integrated circuits

would double every year, and "Moore's Law" has held true for the past 20 years [1].

However, the technology to place billions of transistors in small electronic circuits is currently facing fundamental limits [2]. The continued thinning of silicon for miniaturizing circuits will destroy the band structure for the movement of electrons. These limitations can be surpassed by engineering single molecules to serve as transistors, diodes, and molecular wires. However at the nanoscale quantum effects must be considered in the design of electronic components.

The study of electron transport using organic molecules anchored in electrodes will allow the design of new electronic devices with various functionalities such as: memories, transistors, diodes, molecular switches, conductors, integrated circuits and organic photovoltaic cells devices. Organic molecules can be potential candidates for electronic devices if they present switching, conduction, rectifying behavior, and also if they have complex nonlinear current-voltage (I-V) characteristics such as negative differential resistance (NDR) [3]. Molecular interruption and NDR behavior can take place in an organic molecule attached to electrodes due the changes in the shape, occupation and energy of molecular orbitals (MOs) [3]. To achieve efficient electron transport and high electronic transmission values, strong electronic coupling among the molecule and electrodes is necessary. This coupling is provided by the energetic alignment of the molecular states with the electrode Fermi energy, significant orbital overlap between the structural units within the molecule and at the molecule-electrodes interfaces (spectral density) [4].

Kim et al. [5] pointed out that new electronic devices can be constructed by using external fields such as electrical and magnetic fields. Since an electric field is a powerful mean to change the energy and shape of molecular orbitals, the study of the effect of electrical fields on molecular orbitals would permit the creation of new electronic devices that can function as molecular switches, diodes, rectifiers, etc.

In the last decade there have been many theoretical studies that have used the NEGF-DFT formalism to study the electron transport of molecules attached to electrodes, since this method has been able to reproduce experimental data [6]. Seminario et al. [7] studied theoretically the current-voltage characteristics of the Au_n-S-(p-C6H4)-S-Au_n (n=1–5) molecular junction connected to bulk gold in excellent agreement with a break junction experiment and with other *ab initio* calculations. Seminario et al. [8] studied the current–voltage characteristics of benzene, naphthalene, and anthracene attached to different types of nanoelectrode conformations using gold atoms as electrodes. They found that conductances were high and tunneling barriers were low when compared to saturated alkanes and unsaturated oligophenylene vinylenes.

Kim et al. [9] studied the electrode effect on a molecular electronic device. They studied different electrodes composed of Au, Ru, and carbon nanotubes (CNT) type (5,5) connected by thiolate groups and amide groups to an alkyne molecule using NEGF combined with DFT. Their study showed that the effect of different electrodes and different linkers was noticed by displaying different broadening extent for the given molecular energy levels and different alignment of the energy levels with respect to the Fermi energy of the electrodes.

They obtained nonlinear current-voltage characteristics by using carbon nanotubes as electrodes. Their "CNT-amide" case showed the effect of the zero transmission gap, where the current is almost zero until it reaches 1.5 V, where the broadened LUMO orbital appears in the bias energy windows and the current surges thereafter. It is the LUMO orbital of the alkyne which contributes to the current surge since it is the orbital closer to the Fermi energy of the electrodes.

Lee et al. [10] studied the molecular level alignments and electron transport characteristics based on the NEGF approach combined with DFT for six conjugated molecules (p-terphenyl PTP, and its derivatives, CPTP, NiPTP, CoPTP4, CoPTP5, and FePTP) containing different types of conjugated frameworks. They observe that by combining non and antiaromatic components there was an increase in the conductance since the HOMO level lay closer to the Fermi level, concluding that the manipulation of the conjugated framework is an important factor for controlling the electron transport characteristics.

Car et al. [11] studied the electronic properties of metal–molecule–metal systems at zero bias for benzene–, dibenzene–, and xylyl–dithiol molecules using *ab initio* calculations. Their study focused in understanding how the molecular orbitals change upon adsorption on the electrode and the orbitals closer to the Fermi level, which are important for the electron transport process. For the dithiol-benzene molecular junction, they noticed that the LUMO state is the one closer to the Fermi energy of the electrodes and it is located at 0.16 eV below the Fermi energy. By studying the local density of states (LDOS), they found the existence of states available for conduction around the Fermi energy which are delocalized throughout the molecule. They found that the insertion of methyl group between the sulfur and the carbon ring decouples the two subsystems and suppresses the LDOS at the Fermi energy.

Dilabio et al. [12] carried density functional theory calculations combined with NEGF techniques to model the electron transport through disubstituted benzenedithiol molecules with electron acceptor and electron withdrawing groups. The molecules were bonded by thiol groups to

leads composed of 3x3 gold, 5x5 gold and 3x3 aluminum electrodes. For the disubstituted 3x3 Au-benzenedithiol-Au systems, the absence of transmission states from −0.4 to −0.2 eV, resulted in negative differential resistance in the I-V curve. This NDR feature is not observed in the 5x5 gold electrodes due to the existence of states in the HOMO-LUMO transmission gap. The effect of electron donating and withdrawing groups attached to the benzene was an increase and a decrease in the current, respectively.

Xia et al. [13] used the NEGF formalism combined with density functional theory, to study the electronic transport properties of the butadienimine-based optical molecular switch with and without substituents. The two forms, namely the oxo-amine(keto) and hydroxyl-imine(enol) form, can reversibly switch from each other, making them good candidates for light-driven molecular switches. From the calculated I-V curves they found that the current from the enol form is higher than for the keto form, making it possible to create a molecular switch. By studying the spatial distribution of the frontier orbitals, they found that both the HOMO and the LUMO are delocalized transport channels for the enol, while the HOMO and LUMO for the keto form are both localized orbitals obstructing the electron transport, because the electrons that enter the molecule at the energy of these orbitals have low probability of reaching the other end of the molecule. They also found that the switching performance can be improved by using appropriate electron donors or electron acceptor substituents.

Yang et al. [14] showed the effect that size of the basis set plays in determining the position of the Fermi energy of the lead in the molecular HOMO-LUMO gap, since this position is determined by the molecule-electrode charge transfer. For example: for the Au(111)-OPE3-Au(111) molecular junction, the use of single zeta polarized basis sets leads to the incorrect situation of the Fermi energy being closer to the LUMO state, while using double and triple zeta polarized basis sets gives the Fermi energy closer to the HOMO orbital.

Chelikowsky et al. [15] recently performed first-principles calculations on the electron transport in nanotube junctions decorated with small Au nanoparticles. Their calculations showed that the conductance of nanotube junctions was increased only by the introduction of odd-numbered Au nanoparticles.

Goddard et al. [16] performed NEGF calculations showing that molecular junctions of cyclic and acyclic 1,2-dithiolanes sandwiched between two gold electrodes exhibited essentially the same insulating current–voltage characteristics at moderate bias voltages, despite the difference in their adsorption states.

Zhao et al. [17] performed DFT theoretical calculation about the substituent effect on the molecular wire, oligo (phenylene ethynylene) (OPE), using the electron-donating NH_2 group, and the electron-withdrawing NO_2 group. They considered in the DFT calculations the influence from an external electric field (EF), without the molecules being connected to electrodes and without terminal groups such as thiols, amides, etc. They stated that the electrical properties of molecular electronic devices are related to the electronic structure of the bridge molecule.

Their study focused on the spatial distributions of molecular orbitals, especially frontier orbitals: HOMO and LUMO, as good indicators of electron transport of the molecular wire, only in the absence of quantum interference [18], since frontier orbitals generally dominate the coupling to the electrode and the electron transmission tunnel. Energy levels such as frontier orbitals can be transmission channels [19], stating that a conducting channel is a molecular orbital that is fully delocalized along the molecular backbone and a non-conducting channel is a localized molecular orbital, that cannot connect both ends of the molecule to the metallic electrodes. The important conclusion from their work was the development of a static treatment at the DFT level for predicting the current–voltage behavior, by analyzing only the spatial distribution and energy levels under the influence of an external EF without having to resort to the complex NEGF formalism. Later Zhao et al. [20] investigated molecular rectification in the porphyrin-based molecular junction originating from asymmetrical electrode-molecule contacts using density functional theory combined with the NEGF method. They employed two thiol groups on the left side of the gold surface and one thiol group on the right side to increase the electronic coupling. The Au(111) surface was simulated using a 4×4 cell with periodic boundary conditions, and the supercell consisted of two layers of Au atoms to the left and three layers to the right with 16 atoms for each layer in the scattering region. Their transportation calculations in these asymmetrical junctions showed obvious rectification, an important effect in the design of molecular diodes.

Stokbro et al. [4] by calculating theoretically using NEGF-DFT the conductance of junctions containing molecules connected by thiol end groups, with different levels of conjugation and of different lengths, found that the availability of orbital density on the molecule end contact atoms is an important factor for obtaining high conductance values and can dominate both the conjugation of the molecule and its alignment with the electrode Fermi energy. When the conductive frontier molecular orbital lacks substantial orbital density on the end group, the conduction will be reduced significantly, even in there exists good the alignment between the molecular orbitals and the electrode Fermi energy.

In this work we study the electronic transport of the S-FNPPB-o molecule connected with thiolate groups at the ends as shown in Fig. 1, using two approaches. The first of these is the statical approach proposed by Zhao et al., [17, 20], in which we study the molecular geometric and electronic properties of the molecule under an applied electric field with the thiolate ends fixed. From this study we hope to predict junction properties using the path continuity approach for the electron transport, and a second dynamical approach in which we employ the NEGF-DFT formalism to obtain the current–voltage characteristics at a bias range form −2 to 2 V and the influence of the bias on the transmission spectrum. The objective of this study is to elucidate the electronic transport of S-FNPPB-o to determine if NDR is present in this molecule or note other nonlinear current voltage characteristics that can help in the design of new nanoelectronic devices.

We focus on the S-FNPPB-o fluorinated oligo(phenylene ethynylene)s (OPE) molecule, since this molecule has been previously synthesized by Tour et al. [21] with the aim of using fluorination to alter the thermal, chemical, and electronic properties of OPE molecules and make them suitable for integration with solid state devices. The 4-(3-nitro-4-pentafluorophenylethynyl,phenylethynyl)benzenethiol (FNPPB-o) molecule can form stable self-assembled monolayers (SAMs) in gold and platinum surfaces. Tour et al. also observed that the use of fluorocarbons produced materials with increased thermal stability and chemical resistance, necessary characteristics for high-temperature production processes of SAMs on metal surfaces, such as gas-phase physical vapor deposition (PVD). Tour et al. also stated that molecular electronic devices based on this fluorinated molecule may lower the HOMO energy and the corresponding Schottky energy barrier, increasing the electron transmission from a bulk contact through the frontier orbitals of a chemisorbed organic film.

Tour et al. [22] also observed that FNPPB-o molecules embedded in amide-containing alkanethiol SAMs on Au {111}, presented bias-dependent switching as a function of the interaction between the dipole moment of the OPEs and the electric field applied with the scanning tunneling microscope.
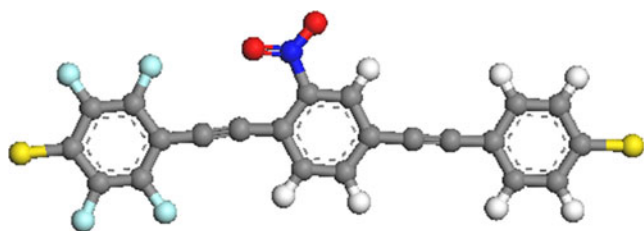
Fig. 1 Schematic representation of the S-FNPPB-o molecule

Computational methodology

Figure 2a shows a schematic representation of the S-FNPPB-o molecular junction and Fig. 2b shows the S-FNPPB-o molecular junction, replacing the fluor atoms with hydrogen atoms named OPE-NO$_2$ molecular junction. Before the introduction of an electrical field EF, the S-FNPPB-o molecule was fully optimized at the GGA-PBE/DNP level of theory using DMOL3 [23–25]. Then, the terminal sulfur atoms were fixed in space to simulate the connection to the electrodes. Then we carry geometrical optimizations at the same level of theory in the presence of an external EF in the neutral state for this molecule since the remaining bonds in the sulfur atoms in the diradical are saturated by gold atoms. Based on experimental results from SAMs attached to gold surfaces, it is generally accepted that hydrogen atoms of the thiol terminal group desorbs, while the sulfur atom in the radical binds strongly to the gold surface [4, 11, 26]. The EF ranging from zero to 0.257 V/Å is defined as uniform and aligned along the two terminal carbon–sulfur inter-atomic vector, comparable with most experimental measurements [17]. Both positive and negative EF where applied to the molecule.

The disadvantage of this static computational methodology based in applying an EF to the S-FNPPB-o molecule is that it does not take into consideration the effect of the electrodes interface in the electronic transport. A more rigorous dynamic study was carried using the NEGF-DFT formalism to obtain the current–voltage characteristics of the molecular junction as well as the transmission spectrum of the molecular junction at different voltages (S-FNPPB-o attached to two Au(111) electrodes).

An overview of the NEGF-DFT formalism has been given by Stokbro et al. [27–29]. The conduction pathway for coherent motion of charge is comprised of three segments: the intramolecular segment and the two molecule-electrode interfaces. To achieve efficient conductance, strong electronic coupling among the three segments should be achieved. Such coupling is provided by close energetic lineup of the molecular states with the electrode Fermi energy (small injection gap) and by significant orbital overlap between the structural units within the molecule and at the molecule-electrodes interfaces (spectral density). We used the Atomistix Toolkit (ATK) software package [30] to study the electronic transport properties of a molecule assembled in two Au (111) electrodes. This program combines DFT with NEGF to simulate the electronic transport in single molecule devices under nonequilibrium conditions.

For the simulation of the molecular junction, the previously optimized S-FNPPB-o molecule at the DFT GGA-PBE/DNP level was translated into the Au junction with the Au(111) surface as a common feature in the
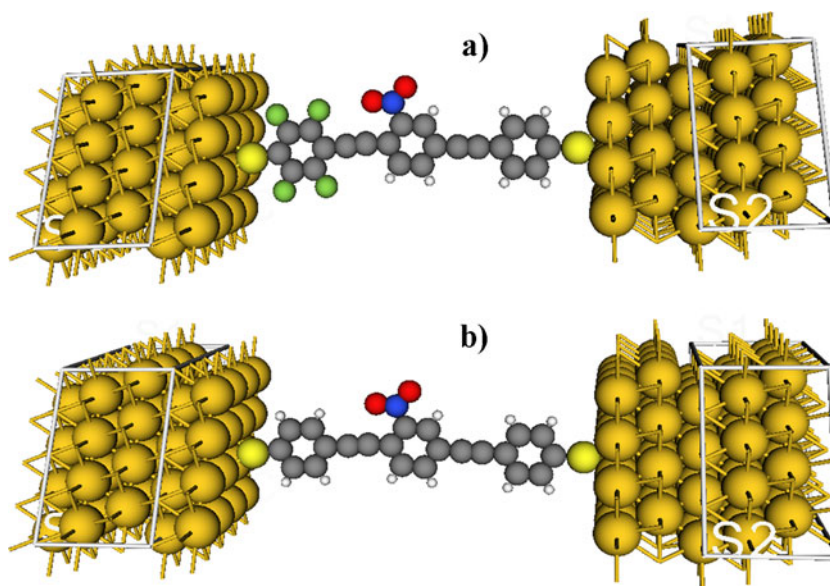
software. (DNP) stands for double numerical plus polarization basis set. The Au(111) surface was simulated using a 4×4 cell with periodic boundary conditions, and the supercell consisted of two layers of gold atoms to the left and three layers to the right with 16 atoms for each layer in the scattering region. The complete molecular junction was relaxed using DNP basis set for all atoms and the GGA-PBE functional in the calculation, and the introduction of relativistic effects for some core electrons. All atoms except the gold atoms, which were kept fixed at their bulk positions, were relaxed until the force on each was less than 0.05 eV/Å. With this procedure we obtained end sulfur-surface distances of 1.95, and 1.97Å in the local energy minimum, that is, the hollow site on the Au surface. This Au-S distance is in agreement with values obtained in the literature [4, 10, 11, 13, 26, 31].

A bias voltage from −2.0 to 2.0 V was applied between the two metal leads, and the electric current was generated using the Landauer-Buttiker formula [32–35].

The current through a single molecule connected by two metal electrodes is given by the following expression:

$$I = \frac{2e}{h} \int\limits_{\mu_R}^{\mu_L} T(E, V_b)[f_R(E, V_b) - f_L(E, V_b)]dE \qquad (1)$$

where $f_R(E, V_b)$, and $f_L(E, V_b)$ are the Fermi-Dirac functions for the left and right electrodes at energy E under the bias voltage, V_b; $T(E, V_b)$ is the sum of the transmission probabilities of all channels available at energy E under bias V_b [36], μ_L and μ_R are the chemical potentials of the left and right electrodes; and $[\mu_L(V_b), \mu_R(V_b)]$ is the energy region for the current referred as the bias energy window. The bias energy window is given by $\mu_L=E_F - eV_b/2$ and

$\mu_R=E_F+eV_b/2$, and E_F is the Fermi energy which is the average value of the chemical potential of the left and right electrodes, is usually set to zero [20].

From this equation 1, it can be observed that only electrons with energies close to the Fermi level and within the bias energy windows add to the current. Since the current is the integral of the transmission coefficient in the bias energy window, analysis of the transmission can provide a comprehensible understanding of the electron transport behavior.

Results and discussions

Table 1 shows the spatial distribution of the HOMO and LUMO for the S-FNPPB-o molecule at different electrical fields, frontier orbitals HOMO and LUMO, can become good indicators of electron transport of the molecular wire in the absence of quantum interference [18] because they dominate the coupling to the electrode and form the electron transmission tunnel, and in the majority of the cases these orbitals align closer with the Fermi energy of the electrodes, becoming important channels for electron transport. In the coherent tunneling regime, the electron transport of a molecular junction can be studied in terms of the continuity of the conduction pathway.

In order to study and be able to modulate the electrical properties of molecular wires, it is important to visualize how the frontiers molecular orbital spatiate in response to an external EF. Taking into consideration that a conducting channel is a molecular orbital that is fully delocalized along the molecular backbone; conversely, a non-conducting channel is a localized molecular orbital, which cannot connect both ends of the molecule to the metallic contacts

Table 1 Spatial distribution of the HOMO and LUMO orbitals under an electric field for the S-FNPPB-o molecule

Electric Field		homo	lumo
ua	V/Å		
-0.005	-0.257		
-0.004	-0.206		
-0.003	-0.154		
-0.002	-0.103		
-0.001	-0.051		
0	0.000		
0.001	0.051		
0.002	0.103		
0.003	0.154		
0.004	0.206		
0.005	0.257		

and the electrons that enter at the energy of these localized orbital have very low probabilities of transmission to the other end. [13, 17, 19].

From Table 1, we can observe that under the EF applied, the HOMO spatial distribution remains localized around only one of the sulfur end atom at most EF, and the HOMO orbital delocalizes only at a negative EF value equal to −0.005 au; on the contrary, we observe that the LUMO orbital delocalizes in the molecular backbone at all positive EF with a large orbital density at the sulfur atoms. At negative EF we observe scarce orbital density at the sulfur

left end side in the EF range from −0.001 to −0.002 au, signaling low electron transport at these EF. From the results from Table 1 we can perceive that if the frontier orbitals align closer to the Fermi energy, the electron transport would go mostly through the LUMO orbital, being the transport larger at positive EF, than at negative EF due to the small orbital density at the left end side on the sulfur atom in the LUMO orbital.

Figure 3a shows the transmission spectrum and density states at 0 V bias for the S-FNPPB-o molecular junction, the zero energy corresponds to the Fermi energy of the electrodes,

Fig. 3 (a) Transmission spectrum and DOS at 0 V for OPE-NO$_2$ and S-FNPPB-o molecular junctions (b) Transmission spectrum and DOS at 2 V for OPE-NO$_2$ and S-FNPPB-o junctions (c) Transmission spectrum and DOS at −2 V for OPE-NO$_2$ and S-FNPPB-o molecular junctions
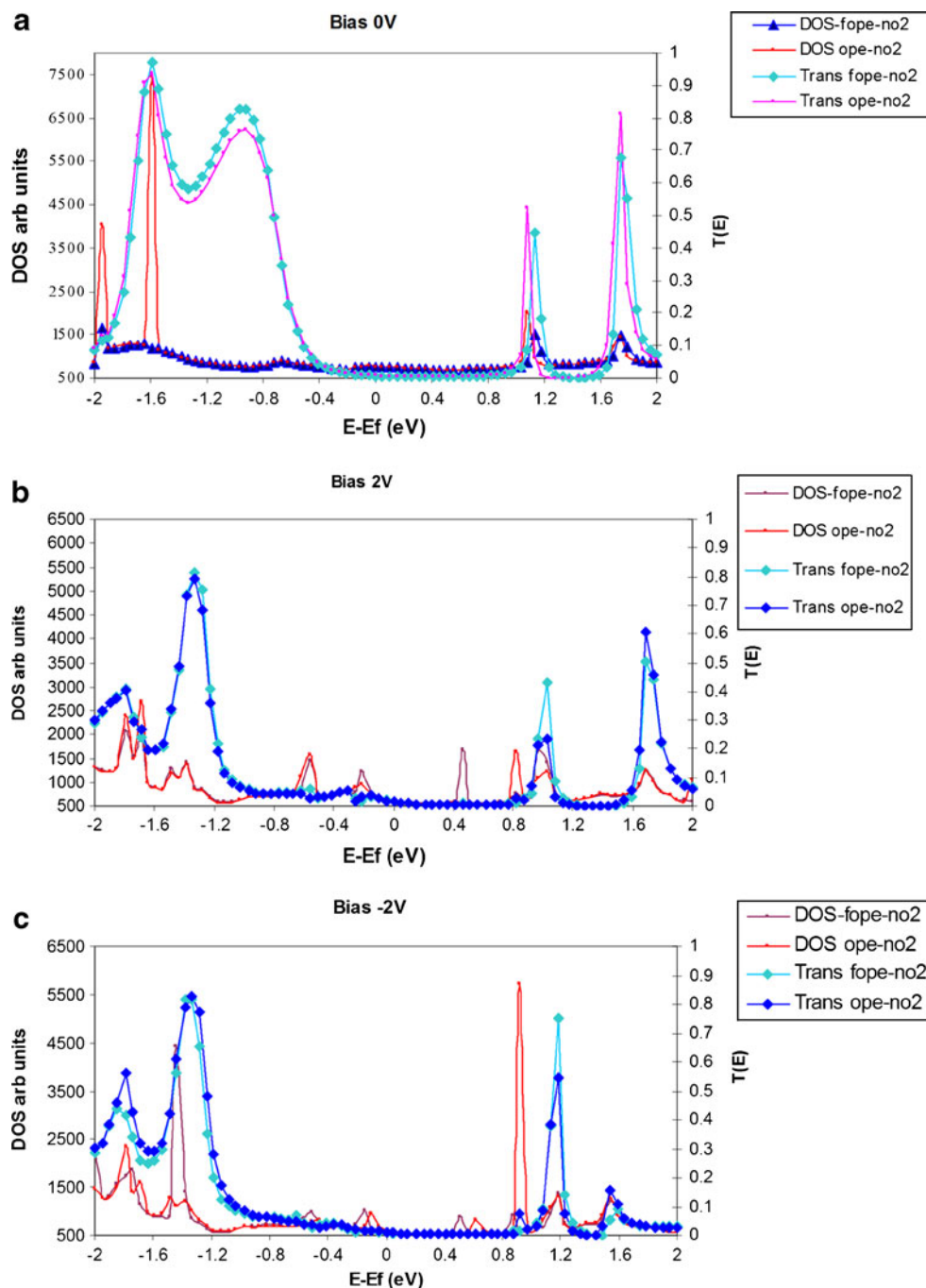
Fig. 4 Complete dependence of the transmission spectrums as a function of the applied voltage range [−2 V, 2 V] for the S-FNPPB-o molecular junction
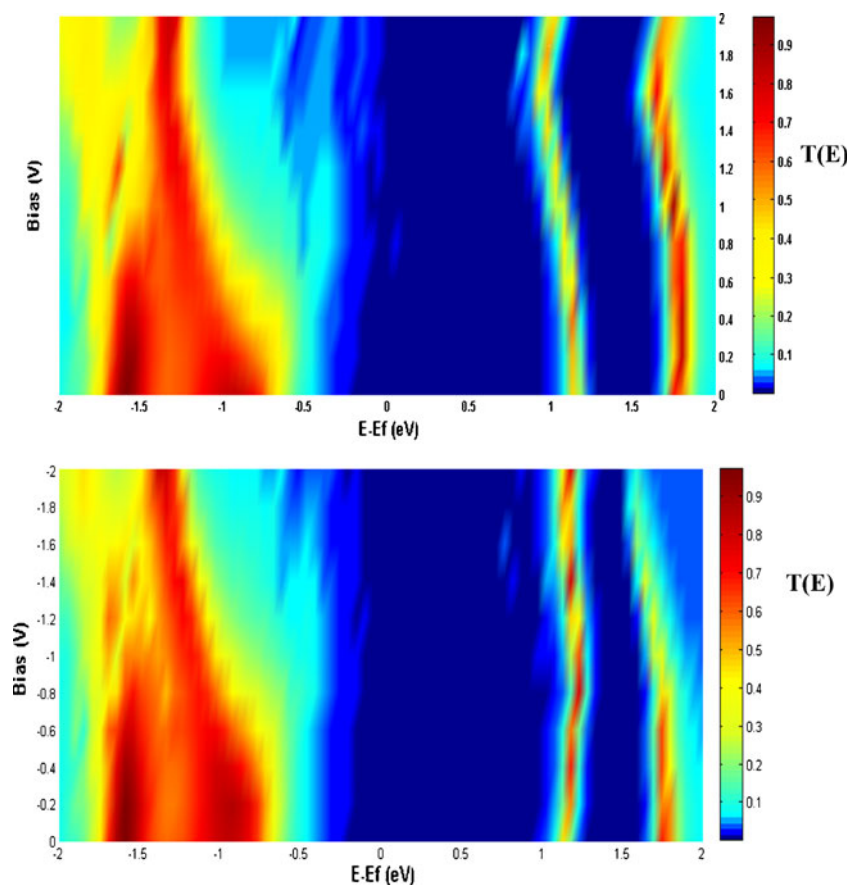
we observe that the LUMO orbital (−0.93 eV) lies much closer to the Fermi level than the HOMO orbital (−1.46 eV), and the electron transmission goes through the unoccupied molecular states, also the LUMO+1 orbital lies at 1.13 eV. From this figure we can see how the LUMO orbital is broader
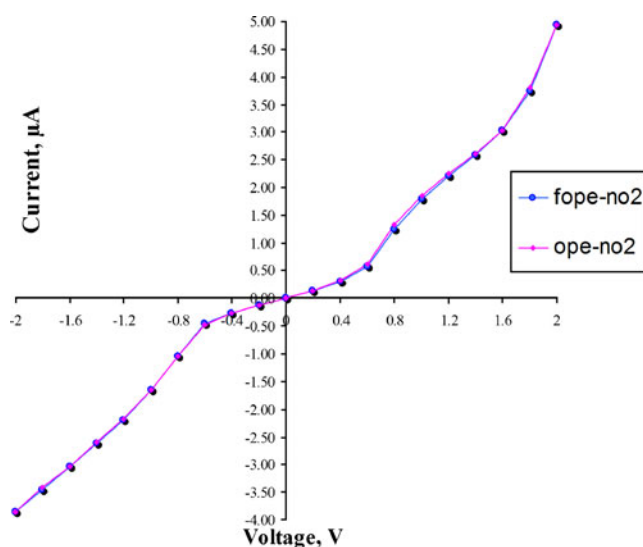
Fig. 5 I-V curves for the OPE-NO$_2$ and S-FNPPB-o molecular junctions

with respect to the HOMO and LUMO+1 orbitals. From Fig. 3a at 0 V we can see a region from 0 to 0.8 eV showing near zero transmission values. This behavior might be due to the lack of states in the electrodes to couple with the molecular orbitals from the S-FNPPB-o bridge.

From the density of states at 0 V in Fig. 3a we can observe a marked peak at the LUMO+1 orbital. In the same Fig. 3a, we compare the DOS and the transmission spectrum at 0 V, for the S-FNPPB-o molecule replacing the fluor atoms with hydrogen atoms named OPE-NO$_2$ molecular junction, in order to assess the effect of the fluorination on the electron transport. We observe only minor differences in the transmission spectrum by comparing both molecular junctions. With respect to the DOS, we observe in the OPE-NO$_2$ molecular junction the appearance of a marked peak near the HOMO orbital, absent in the S-FNPPB-o molecular junction.

The bias dependence of the transmission spectrum and DOS at 2 V and at -2 V is shown in Figs. 3b and c. From these figures we can see that the LUMO+1 orbital becomes closer to the Fermi energy at 2 V than at -2 V. We can also observe that the broad LUMO peak at 0 V center at −0.9 eV does not appears at this energy at -2 V and 2 V, and at these voltages there are strong peaks now centered at −1.33 eV and −1.79 eV.
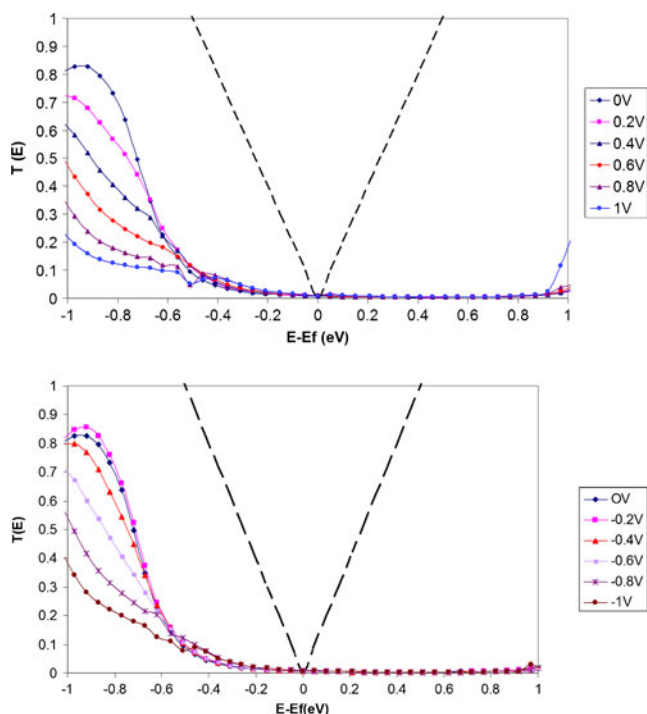
Fig. 6 Transmission spectrums in the bias energy window from -1 eV to 1 eV for voltages from 0 to ±1 V for the S-FNPPB-o molecular junction. Dotted black lines indicate the range of current integration (bias energy window) around the Fermi level

Figure 4a shows the complete bias dependence of the transmission spectrum from 0 to 2 V voltages and Fig. 4b shows the complete bias dependence of the transmission spectrum from 0 to -2 V voltages. Both graphs display similar patterns except for the behavior of the LUMO+1 band and different transmission states around −0.4 eV.

Since the current is the integral of the transmission coefficient within the bias window around the Fermi level, we observe in Fig. 5 very low values for the current at low bias from 0 to ±0.8 V since the small transmission values that enter the bias window are due to the tails of the

broad, LUMO orbital, higher current values at positive bias than at negative bias. As we exceed ±0.8 V from Fig. 6 in the corresponding bias window [−0.4 eV, 0.4ev] we can observe higher transmission states from the broad LUMO orbital that enter into the bias energy window arising from −0.2 eV till −0.4 eV with transmission values from 0.05 till 0.1 and we observe a steep slope on the transmission from these states and the current increases considerably being a little higher at positive bias than at negative bias, since the transmission values are higher at positive bias than at negative bias. When the voltage approaches ±2 V the LUMO+1 orbital enters into the Bias energy window and the current surges again, more at the 2 V positive bias.

It is interesting that we also calculated the I-V curve for the OPE-NO$_2$ molecular junction and compared with our fluorinated molecular junction, the current changes slightly over the whole range by less than 5% percent. This results is in agreement with the calculations of Stokbro et al. [37] that carried NEGF-DFT calculations to determine the I-V curve for the same OPE-NO$_2$ molecular junction attached to Au(111) electrodes, in the absence of NO$_2$ functionalization, with an NO$_2$ group, and with an NH$_2$ group instead of the NO$_2$ group, and found almost no effect on the current values under the [−2 V, 2 V] voltage range. They also noticed that around ±1.8 V there was an abrupt increase in the current.

Figure 7, shows the energy spectrum of the molecular projected self-consistent Hamiltonian (MPSH) onto the S-FNPPB-o molecular orbitals, the Fermi level of the gold electrode is set to be zero. What is meant with MPSH, is that the self-consistent Hamiltonian is projected onto the atoms of the S-FNPPB-o molecule, and then diagonalized to produce an energy spectrum. This information shows how the molecular levels for the S-FNPPB-o molecule are affected when placed between the two gold electrodes under the influence of a voltage [30]. The alignment of the

Fig. 7 HOMO and LUMO energy levels as functions of the applied bias for the OPE-NO$_2$ and S-FNPPB-o molecular junctions
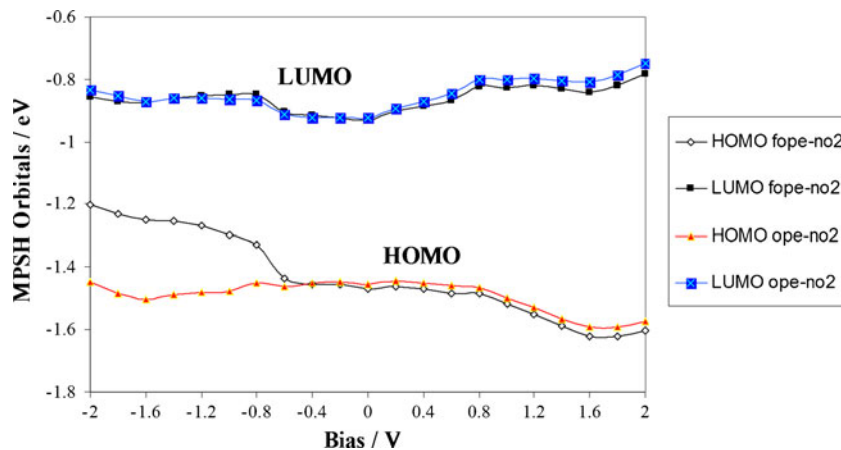
molecular orbitals in relation to the Fermi level is very important, since states near the Fermi level are the main contributors to electron transport [36].

From Fig. 7 we can observe that the HOMO for the S-FNPPB-o molecular junction decreases under positive bias and increases at negative bias, while the LUMO for the S-FNPPB-o molecular junction, remains more or less at the same level, leading to an asymmetrical HOMO-LUMO gap (HLG). Compared with the OPE-NO_2 molecular junction, the effect of the fluorination is to increase the HOMO orbital from −0.8 to −2 V. At other voltages the HOMO levels for both junctions do not differ significantly. Also, when comparing the LUMO orbitals for both junctions we do not observe any significant difference.

Conclusions

We have presented theoretical calculation where an electric field was applied to the S-FNPPB-o molecule without gold electrodes, 'static treatment'. From the visualization of the spatial distribution of the frontier orbitals, we noticed that only the LUMO orbital is delocalized at the molecular backbone at most EF ranges, signaling and important electron transport channel. We perceive qualitatively that the electron transport would be larger at positive EFs than at negative EFs, due to the small orbital density at the left end side on the sulfur atom at some negative EFs values in the LUMO orbital.

By carrying NEGF-DFT calculations we calculated the I-V curve for the S-FNPPB-o molecular junction and the bias dependence of the transmission spectrum. The transmission spectrum at 0 Volts has three peaks closer to the Fermi energy of the electrodes. However, only one transmission peak corresponding to the LUMO orbital is closer to the Fermi energy. That means that there are three electron transport channels, but only one of the channels, namely the LUMO, makes its contribution to the current in the [−0.9 eV, 0.9 eV] energy bias window. Only when the voltage approaches ±2 V does the LUMO+1 orbital enter into the [−1 eV, 1 eV] bias energy window and we observe an increase in the current, but at the voltage range studied the HOMO orbital never enters into the bias energy window.

From our calculations we observed that the I-V curve for the OPE-NO_2 molecular junction compared with our S-FNPPB-o fluorinated molecular junction changes by less than 5% percent, so the fluorination does not change the I-V characteristics in the OPE-NO_2 molecular junction in the voltage range studied.

It is interesting to note that both first principle calculations, the static EF treatment and the NEGF-DFT formalism, coincide in that the electronic transport has to be through the unoccupied frontier orbitals.

References

1. Moore GE (1965) Cramming more components onto integrated circuits. Electronics 8:114–117
2. Service RF (2001) Molecules get wired. Science 294:2442–2443. doi:10.1126/science.294.5551.2442
3. Seminario JM, Araujo RA, Yan LJ (2004) Negative differential resistance in metallic and semiconducting clusters. J Phys Chem B 108:6915–6918. doi:10.1021/jp0377811
4. Revital C, Stokbro K, Martin JML, Ratner MA (2007) Charge transport in conjugated aromatic molecular junctions: Molecular conjugation and molecule − electrode Coupling. J Phys Chem C 111:14893–14902. doi:10.1021/jp0795309
5. Kim WY, Kim KS (2010) Tuning molecular orbitals in molecular electronics and spintronics. Acc Chem Res 43:111–120. doi:10.1021/ar900156u
6. Lindsay SM, Ratner MA (2007) Molecular transport junctions: Clearing mists. Adv Mater 19:23–31. doi:10.1002/adma.200601140
7. Derosa PA, Seminario JM (2001) Electron transport through single molecules: scattering treatment using density functional and green function theories. J Phys Chem B 105:471–481. doi:10.1021/jp003033+
8. Yan L, Bautista EJ, Seminario JM (2007) *Ab Initio* analysis of electron currents through benzene, naphthalene, and anthracene nanojunctions. Nanotechnology 18:485701(1–8). doi:10.1088/0957-4484/18/485701
9. Cho Y, Kim WY, Kim KS (2009) Effect of electrodes on electronic transport of molecular electronic devices. J Phys Chem A 113:4100–4104. doi:0.1021/jp810467q
10. Lee SU, Belosludov RV, Mizuseki H, Kawazoe Y (2007) Control of electron transport by manipulating the conjugated framework. J Phys Chem C 111:15397–15403. doi:10.1021/jp074294n
11. Piccinin S, Selloni A, Scandolo S, Car R, Scoles G (2003) Electronic properties of metal-molecule-metal systems at zero bias: A periodic density functional study. J Chem Phys 119:6729–6735. doi:10.1063/1.1602057
12. Smeu M, Wolkow RA, DiLabio GA (2008) Theoretical investigation of electron transport modulation through benzenedithiol by substituent groups. J Chem Phys 129:034707(1–8). doi:10.1063/1.2955463
13. Xia C, Liu D, Fang C, Zhao P (2010) The I-V characteristics of the butadienimine-based optical molecular switch: An *ab initio*. Physica E 42:1763–1768. doi:10.1016/j.physe.2010.01.044
14. Ke S, Baranger HU, Yang W (2007) Electron transport through single conjugated organic molecules: Basis set effects in *ab initio* calculations. J Chem Phys 127:144107-1-144107-6. doi: 10.1063/1.2770718
15. Khoo KH, Chelikowsky JR (2009) Electron transport across carbon nanotube junctions decorated with Au nanoparticles: Density functional calculations. Physical review B 79:205422(1–6). doi:10.1103/PhysRevB.79.205422
16. Jang YH, Goddard WA III (2008) Electron transport through cyclic disulfide molecular junctions with two different adsorption states at the contact: A density functional theory study. J Phys Chem C 112:8715–8720. doi:10.1021/jp800201z
17. Li Y, Zhao J, Yin G (2007) Theoretical investigations of oligo (phenylene ethylene) molecular wire: Effects from substituents

and external electric field. Comput Mater Sci 39:775–781. doi:10.1016/j.commatsci.2006.09.010

18. Solomon GC, Andrews DQ, van Duyne RP, Ratner MA (2008) When things are not as they seem: Quantum interference turns molecular electron transfer "rules" upside down. J Am Chem Soc 130:7788–7789. doi:10.1021/ja801379b

19. Nitzan A, Ratner MA (2003) Electron transport in molecular wire junctions. Science 30:1384–1389. doi:10.1126/science.1081572

20. Zhao J, Yu C, Wang N, Liu H (2010) Molecular rectification based on asymmetrical molecule-electrode contact. J Phys Chem C 114:4135–4141. doi:10.1021/jp905713a

21. Maya F, Chanteau SH, Cheng L, Stewart MP, Tour JM (2005) Synthesis of fluorinated oligomers toward physical vapor deposition molecular electronics candidates. Chem Mater 17:1331–1345. doi:10.1021/cm0486161

22. Lewis PA, Inman CE, Maya F, Tour JM, Hutchison JE, Weiss PS (2005) Molecular engineering of the polarity and interactions of molecular electronic switches. J Am Chem Soc 127:17421–17426. doi:10.1021/ja055787d

23. a) www.accelrys.com

24. Delley B (1990) An all-electron numerical method for solving the local density functional for polyatomic molecules. J Chem Phys 92:508–518. doi:10.1063/1.458452

25. Delley B (2000) From molecules to solids with the DMol3 approach. J Chem Phys 113:7756–7764. doi:10.1063/1.1316015

26. Xia CJ, Fang CF, Zhao P, Liu HC (2010) Effects of contact atomic structure on the electron transport of pyridine-substituted dithienylethene optical molecular switch. Eur Phys J D 59:375–378. doi:10.1140/epjd/e2010-00196-2

27. Stokbro K, Taylor J, Brandbyge M, Mozos JL, Ordejon P (2003) Theoretical study of the nonlinear conductance of Di-thiol benzene coupled to Au(1 1 1) surfaces via thiol and thiolate bonds. Comput Mater Sci 27:151–160. doi:10.1016/S0927-0256(02)00439-1

28. Stokbro K (2008) First-principles modeling of electron transport. J Phys Condens Mater 20:064216(1–6). doi:10.1088/0953-8984/20/6/064216

29. Datta S (2005) Quantum Transport: Atom to transistor. Cambridge University Press, New York

30. www.quantumwise.com

31. Miao L, Seminario JM (2007) Electronic and structural properties of oligophenylene ethynylenes on Au(111) surfaces. J Chem Phys 126:184706(1–7). doi:10.1063/1.2734545

32. Datta S (1996) Electronic Transport in the Mesoscopic Systems. Cambridge University Press, New York

33. Meir Y, Wingreen NS (1992) Landauer formula for the current through an interacting electron region. Phys Rev Lett 68:2512–2515. doi:10.1103/PhysRevLett.68.2512

34. Landauer R (1957) Spatial variation of currents and fields due to localized scatterers in metallic conduction. IBM J Res Dev 1:223–231. doi:10.1147/rd.13.0223

35. Buttiker M (1988) Coherent and sequential tunneling in series barriers. IBM J Res Dev 32:63–75. doi:10.1147/rd.321.0063

36. Bai P, Li E, Neerja CP (2005) Theoretical investigation of metal-molecule interface with terminal group. IEEE Trans Nanotechnol 4:422–429. doi:10.1109/TNANO.2005.851282

37. Stokbro K, Taylor J, Brandbyge M, Ordejón P (2003) TranSIESTA: A spice for molecular electronics. Ann NY Acad Sci 1006:212–226. doi:10.1196/annals.1292.014

ORIGINAL PAPER

Polaron binding energy in polymers: poly[methyl(phenyl) silylene]

Juraj Nožár · Stanislav Nešpůrek · Jakub Šebera

Abstract This paper presents a theoretical approach to the evaluation of polaron binding energy in polymers. Quantum chemical calculations were performed on a model polymer, poly[methyl(phenyl)silylene], employing the B3LYP and CAM-B3LYP method. The polaron binding energy consists of two terms: the molecular deformation energy and electron-phonon term. Its value was found to be about 0.23 eV at the CAM-B3LYP/6-31G* level of theory.

Keywords Polaron · Polaron binding energy · Polysilane · Poly[methyl(phenyl)silylene] · Quantum chemical calculations

Introduction

Polysilanes are potential materials for the fabrication of molecular wires, which are essential elements in molecular electronics [1, 2]. Precise knowledge concerning the charge carrier transport in these materials is thus of great importance. The polysilane chain, composed of single-bonded silicon atoms, is σ-conjugated when it is derivatized with the proper side groups and the all-trans conformation has been achieved. This kind of conjugation cannot be found in carbon-backbone polymers, because carbon σ-orbitals do not extend far enough to achieve an efficient overlapping.

A slowly moving charge carrier induces a deformation of a soft macromolecule; the atoms in a charged molecule shift to new positions in order to preserve their minimal potential energy. The new conformation and charge redistribution also result in a shift of the frequencies of vibration modes. All of these variances move along with the charge carrier (together they form the so-called polaron quasiparticle) and create a potential well with the depth being referred to as the polaron binding energy, E_p. This self-trapping process is reflected in a decrease of the charge-carrier mobility; it is thus an essential parameter of the charge-carrier transport phenomenon in molecular materials.

The experimental estimation of the polaron binding energy is not simple. We have developed a method for its determination from the temperature and voltage dependencies of charge-carrier mobility. The principle of the method has been described before [3] and the polaron binding energy was found to be E_p=0.29 eV. Through a linear extrapolation of the temperature dependence of the activation energy of the charge-carrier mobility to zero temperature, Bässler et al. reported a value of E_p=0.16 eV [4]. However, the dependence is not generally linear, and therefore this E_p value is only a rough estimate. The reasoning can also be based on the value of E_p=0.08 eV presented by Pan et al. [5] for excitons. Regarding the fact that in molecular crystals the defects of the same nature form about three times deeper traps for charge carriers than for singlet excitons [6], one can estimate the polaron binding energy as E_p=0.24 eV.

J. Nožár · S. Nešpůrek · J. Šebera
Institute of Macromolecular Chemistry, ASCR,
v.v.i., Heyrovského nám. 2,
162 06 Prague 6, Czech Republic

J. Nožár (✉)
Charles University in Prague, Faculty of Mathematics and Physics,
Ke Karlovu 3,
121 16 Prague 2, Czech Republic
e-mail: jurajnozar@seznam.cz

S. Nešpůrek
Faculty of Chemistry, Brno University of Technology,
Purkyňova 118,
612 00 Brno, Czech Republic

Detailed studies of the conformational and population analysis in poly[methyl(phenyl)silylene] (PMPSi) have been provided by Toman et al. [7, 8]. They have shown that the introduction of an excess charge into the PMPSi chain significantly alters the geometry of the backbone. They used a hybrid B3LYP [9–11] method, which has recently been improved in the form of the CAM-B3LYP method (the Coulomb-attenuating method) [12]; the B3LYP method uses a constant ratio of 0.19 for the Hartree-Fock (HF) and 0.81 for the Becke 1988 (B88) exchange functional for any spacing between interacting systems. The CAM-B3LYP method uses the same ratio for the short-range interactions, but with increasing interaction distance it shifts the ratio in favor of the HF term through a smooth standard error function with a parameter of 0.33. At a long range, the exchange potential is described by 0.65 HF plus 0.35 B88.

In this paper, we will present the utilization of quantum chemical calculations for the determination of the energetics of system relaxation in the presence of an excess charge carrier in model silane polymers.

Theoretical approach

Figure 1 shows a schematic diagram of the potential energy curves of a molecule with an excess positive charge $((M_0h)^+$, i.e., a molecule in the so-called *hot* state) and a positive molecular polaron (M^+, a molecule in the *relaxed* state). The curve of the potential energy of $(M_0h)^+$ has a minimum at the q_0 coordinate (q represents the atom's position in the molecule) with the corresponding vibration frequency ω_0, whereas the position of the potential minimum of M^+ has shifted to q_p and the vibration frequency has changed to ω_p. The difference between the potential energy minima, including the zero-point vibrations, represents the polaron binding energy E_p for a single vibration mode. According to our model, the polaron

Scheme 1 The chemical structure of poly[methyl(phenyl)silylene] (PMPSi)

binding energy consists of two terms; (1) the deformation energy of the molecule (E_{def}) and (2) the electron-phonon (vibration) energy (E_{e-ph}):

$$E_p = E_{def} + E_{e-ph}. \tag{1}$$

Both terms can be evaluated by quantum chemical calculations. The deformation energy can be determined from the optimized geometries and the electronic energies of the molecular ion and polaron (see below) and represents the difference between bottom tips of the potential curves in Fig. 1. The energy of the electron-phonon coupling can be calculated according to the equation

$$E_{e-ph} = \frac{1}{2} \sum_{k=1}^{3M-6} \hbar \left(\omega_0^{(k)} - \omega_p^{(k)} \right), \tag{2}$$

where $\hbar$ is the reduced Planck constant, $\omega_0^{(k)}$ and $\omega_p^{(k)}$ are the angular frequencies of the kth vibration mode of the neutral and positively charged molecule (PMPSi is the semiconductor of the P-type), respectively. The summation is performed over all of the vibration modes of the molecule (M is the number of atoms in the molecule). This term represents a correction for the deformation energy to the zero-point vibrations.

Note that in molecular crystals, which have been studied extensively in the past [6, 13], the lattice is more rigid in comparison with the polymer chain and the shift of the equilibrium positions of the atoms is small. Therefore, the main contribution to the polaron binding energy follows from the vibronic term (Eq. 2).

Fig. 1 A schematic energy diagram of the molecular ion $(M_0h)^+$ and molecular polaron (M^+) (see details in the text for an explanation of the symbols)

Table 1 The geometry parameters of the optimized *neutral* MPSi-16-mer obtained at different levels of theory. The following symbols are used in the table: B is the average Si-Si bond length, A is the average angle between the silicon atoms and D is the average dihedral angle between the silicon atoms in the backbone. All of the averages exclude the silicon atoms at the ends of the chain

| Level of theory	B (Å)	A (°)	D (°)
B3LYP/6-31G*	2.389	111.6	173.7
B3LYP/STO-3G*	2.277	112.5	176.2
CAM-B3LYP/6-31G*	2.369	111.3	177.4
CAM-B3LYP/STO-3G*	2.260	111.8	175.3
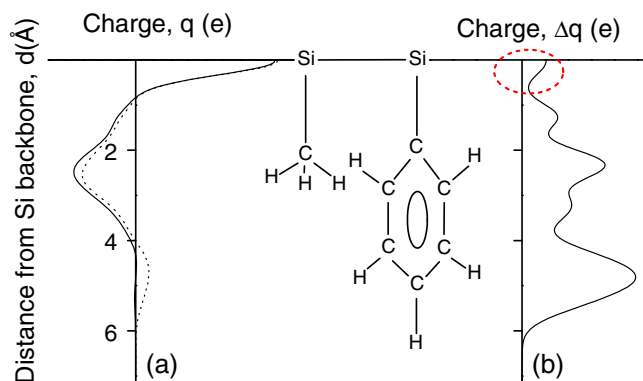
Fig. 2 (a) The charge distribution on the side phenyls and methyl groups of the *neutral* (full line) and *relaxed* (dashed line) state of MPSi-16-mer (calculated at the B3LYP/6-31G* level; similar qualitative results have been obtained at all of the other levels of theory); (b) the difference between the charge distributions on the side groups in the *neutral* and *relaxed* state (the difference of the curves in (a); see the details in the text)

Methods

The Gaussian 09 program package [14] was used for the structure optimization and infrared spectra calculations, employing the B3LYP and CAM-B3LYP methods. Both methods were combined with the 6-31G*[15] and STO-3G* [16] basis sets (see Ref. [7] for the detailed use of the B3LYP method for polysilanes). The latter basis set is rather small, but it allowed us to calculate the vibration spectra of the oligomers longer than eight monomer units and thus to confirm the dependence of the polaron binding energy on chain length. The values obtained from the simulations with this basis set were always taken into account with respect to the results of the calculations with the 6-31G* basis set.

The geometries of the PMPSi (see Scheme 1) molecule and its charged form were optimized for the all-trans conformation, which is assumed to be the ground state geometry [17]; no symmetry assumptions were taken into account. The ends of the chain were capped with methyl groups. The calculations were performed for an isolated molecule in vacuum with no influence of any solution or other surroundings. The

results obtained for the polaron binding energy thus comply with the small polaron approximation [13], where the intermolecular interactions are neglected and only the intramolecular ones are taken into account.

First, the geometry of the neutral molecule (singlet state) was optimized; this state is referred to as *neutral* in the rest of the text. Using this geometry in the cation radical state, the electronic energy of the *hot* state was calculated. The *hot* state is a molecule which is already missing one electron but has not yet relaxed to its ground state. Its electronic energy is labeled as E_{hot} in the text. Afterward, the neutral optimized geometry was taken as the initial geometry and the geometry of cation radical was optimized (using the unrestricted method; all of the optimized molecules yield a spin contaminant equal or very close to the expected value of 0.75). This state is called *relaxed* hereinafter (the geometry and energy of cation radical) and its electronic energy is referred to as $E_{relaxed}$. The infrared spectra were subsequently calculated using the optimized structures (*neutral* and *relaxed*). The deformation energy was estimated using the energies of the *hot* and *relaxed* states according to Eq. 3:

$$E_{def} = E_{hot} - E_{relaxed}. \tag{3}$$

Results and discussion

Optimized geometry parameters

The parameters of the geometries of the MPSi-16-mer optimized at all of the levels of theory are summarized in Table 1. The optimized structures obtained with the 6-31G* basis set exhibit significantly longer Si-Si bonds when compared to the bond lengths with the STO-3G* basis set. The results for the 6-31G* basis set are in good agreement with the experimental measurements, amounting to (2.37 ± 0.02) Å (in poly(di-n-hexylsilylene) determined from EXAFS experiments [18]); the angles and dihedral angles are in better agreement. The differences between the *neutral* and *relaxed* molecule are similar for both basis sets:

Fig. 3 (a) The partial charges on the Si atoms of the backbone in the *relaxed* state; (b) the amount of excess charge localized on the Si atoms, when the molecule transforms from the *neutral* to *relaxed* state. The results obtained at the B3LYP/6-31G* (■), B3LYP/STO-3G* (▲), CAM-B3LYP/6-31G* (□) and CAM-B3LYP/STO-3G* (△) levels on optimized MPSi-16-mers
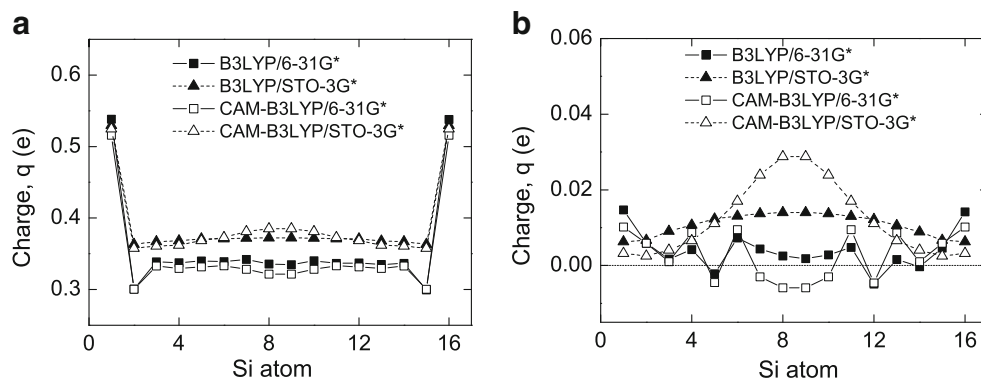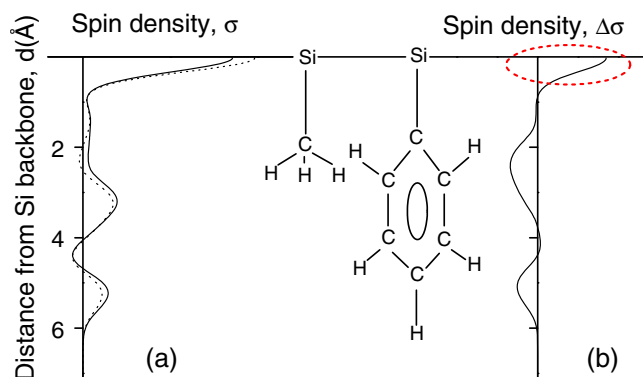
Fig. 4 (**a**) The spin density distribution on the side phenyls and methyl groups in the *hot* (full line) and *relaxed* (dashed line) states of MPSi-16-mer (calculated at the B3LYP/6-31G* level; similar results have been obtained at the other levels of theory as well); (**b**) the difference between the spin-density distributions on the side groups in the *hot* and *relaxed* states (the difference of the curves in (**a**))

Si-bonds in the *relaxed* structure are longer by about 1-2%, the Si angles are smaller by about 5%, and Si-dihedral angles remain about the same (detailed studies using the 6-31G* basis set and on the conformational changes have been published elsewhere [7, 19]).

Charge distribution

The silicon backbone is positively charged even in the *neutral* state while the side phenyls have a partial negative charge. The addition of an excess positive charge results in a charge redistribution during the relaxation process, which has only a minor effect on the Mulliken charges on the Si atoms. Most of the excess charge is localized on the side groups, which arises from the Columbic repulsion. The plots in Fig. 2a represent the charge distribution along the silicon backbone, and the methyl and phenyl side groups. They were obtained by the following procedure: The Si backbone was approximated by an axis connecting the first and the last silicon atom in the chain, after which the perpendicular distances of the atoms in the side groups from this virtual axis were determined and the charges were

summed (e.g., if the p-hydrogens on two different phenyls have the same distance of 6 Å from the Si-axis and one of them has a charge of 0.2e and the other one 0.3e, then the charge in the graph is plotted as 0.5e). When interpreting the curves in Fig. 2a, one must take into account that the values represent the charges obtained by the integration in an angle of 360° around the Si-axis at a certain distance; the different types of atoms, directions from the Si-axis and side groups have not been distinguished between (e.g. if hydrogen on methyl and carbon on phenyl have the same distance from the Si-axis, their charges are summed). Note that the diagram of the molecule in Fig. 2 is only schematic and does not comply with the all-trans conformation of the modeled PMPSi molecule.

As follows from Fig. 2b, most of the excess charge is localized on the side groups. Figure 3a shows the partial charges on the Si atoms in the *relaxed* state for both basis sets. One can see that the use of the STO-3G* basis set leads to a greater concentration of the charge in the central region of the chain as compared with the results obtained with 6-31G*. Figure 3b shows the difference of the charge distributions over the silicon atoms in the backbone between the *neutral* and *relaxed* states. The results using the 6-31G* basis set significantly differ from those obtained with the STO-3G* basis set, because the latter concentrates most of the additional charge to the central region of the backbone. Because of Columbic repulsion, one would expect that the excess charge would be localized mostly at the ends of the chain as follows from the results obtained using the 6-31G* basis set.

Spin distribution

The level of the delocalization of the excess charge was studied by means of the spin distribution over the silicon backbone. When compared to the charge distribution, both basis sets provided similar qualitative and quantitative results. Figure 4a shows the distribution of the spin densities on the side phenyl and methyl groups for the *hot* and *relaxed* molecule of MPSi-16-mer (the details of the

Fig. 5 (**a**) The spin distribution on the Si atoms of MPSi-16-mer in the *relaxed* state; (**b**) the dependence of the total spin density on the Si atoms on the chain length in the *relaxed* state. The results have been obtained at the B3LYP/6-31G* (■), B3LYP/STO-3G* (▲), CAM-B3LYP/6-31G* (□) and CAM-B3LYP/STO-3G* (△) levels
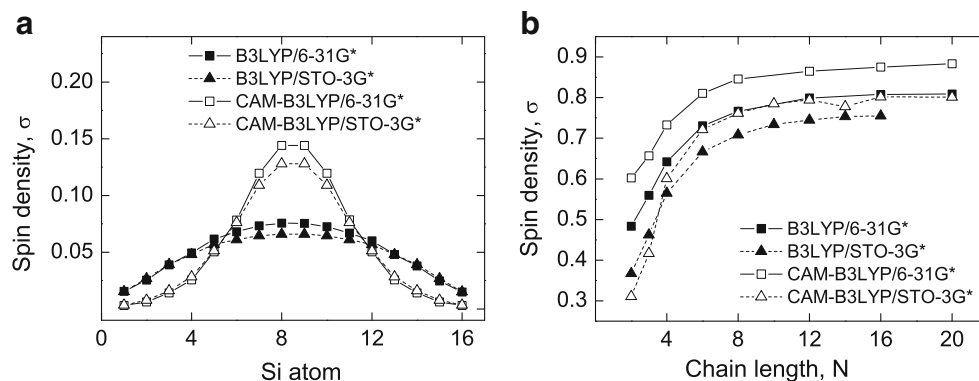
Table 2 Selected vibration frequencies of the experimental (the ATR method) and the theoretical (MPSi-8-mer; the B3LYP method using the 6-31G* and STO-3G* basis sets) infrared spectra. For both basis sets, the *raw* calculated values are shown, from which the average scaling factor (SF) was determined and was subsequently used for the calculation of the "*scaled*" values

Exp.	Theory				Type of vibration
ν (cm^{-1})	6-31G*		STO-3G*		
	ν (cm^{-1})				
	raw	scaled	raw	scaled	
618	642	615	727	634	Si-CH$_3$ valence
668	689	660	749	654	–
696	714	684	774	675	Out-of-plane ring deformation
727	810	776	897	783	C-H deformation on methyl
783	826	791	919	802	C-H deformation on methyl
–	856	820	949	828	C-H deformation on methyl
998	1014	971	1104	963	Planar C-H deformation on phenyl
1025	1119	1072	1158	1010	Planar C-H deformation on phenyl
1098	1125	1078	1230	1073	Si-phenyl valence
1247	–	–	1425	1243	–
1300	1325	1269	1505	1313	Symmetric def. of C-H on methyl
1334	1373	1315	1548	1351	Planar C-H deformation on phenyl
1427	1473	1411	1620	1414	Planar C-H deformation on phenyl
1484	1535	1471	1708	1490	Planar C-H deformation on phenyl
2896	3043	2915	3291	2872	Symmetric valence def. C-H on methyl
2957	3116	2985	3456	3016	Non-symmetric of C-H on methyl
3050	3189	3055	3462	3021	Valence of C-H on phenyl
3068	3205	3071	3273	3030	Valence of C-H on phenyl
SF	0.958[a]	–	0.873	–	

[a] The general value suggested for the B3LYP/6-31G* level of theory is 0.961 [21]

creation of these plots are described above). Figure 4b shows the difference between the spin distribution of the *hot* and the *relaxed* states of MPSi-16-mer. In contrast to the charge distribution, most of the excess spin density is localized on the silicon atoms (the area marked with the ellipsis). Figure 5a depicts the distribution of the spin density over the silicon backbone (a detailed study of this distribution can be found elsewhere [7]). It is clear that the highest spin density is in the middle of the chain, as opposed to the charge distribution (compare with Fig. 3a). Also note that the B3LYP method produces a much broader distribution of the spin density, which is probably caused by the failure of the exchange potential term for long distances. A similar contrast can be found when the B3LYP method is compared to the BH&LYP (with 50% of the HF exchange contribution) spin distribution [20], where it can be seen that a greater HF contribution is necessary for proper long-range interactions. Figure 5b shows the dependences of the total spin densities on the chain length.
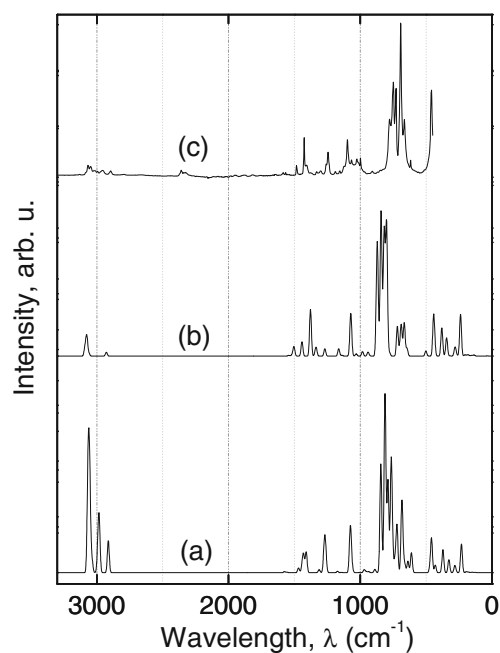
Fig. 6 (**a**) The theoretical spectrum of MPSi-8-mer acquired by the B3LYP/6-31G* method, a scaling factor of 0.958; (**b**) the theoretical spectrum of MPSi-8-mer acquired by the B3LYP/STO-3G* method, a scaling factor of 0.873); (**c**) the experimental IR spectrum of PMPSi (ATR, a range of 3300–450 cm^{-1})

The amount of the spin density on the silicon atoms increases with the length of the polymer chain which suggests an increasing conjugation on the polymer backbone.

Infrared spectra

The calculated infrared spectra are in good agreement with the experimental data (see Table 2; the theoretical frequencies and intensities shown in Fig. 6 were convoluted with a Gaussian shape with a FWHM of 7cm^{-1}). The scaling

Table 3 The calculated values of the deformation energy, E_{def}, for oligomers of different lengths. The following numbers have been used in the table: (1) is the B3LYP/6-31G*, (2) is the B3LYP/STO-3G*, (3) is the CAM-B3LYP/6-31G* and (4) is the CAM-B3LYP/STO-3G*

Length	(1)	(2)	(3)	(4)
2	0.377	0.710	0.596	0.511
3	0.290	0.511	0.416	0.374
4	0.258	0.242	0.388	0.375
6	0.244	0.221	0.357	0.355
8	0.215	0.212	0.328	0.312
12	0.183	0.174	0.286	0.270
16	0.137	0.143	0.261	0.221
20	0.106	0.123	0.249	0.220

Table 4 The calculated values of the electron-phonon term, $E_{e\text{-}ph}$, for oligomers of different lengths. The following numbers have been used in the table: (1) is the B3LYP/6-31G*, (2) is the B3LYP/STO-3G*, (3) is the CAM-B3LYP/6-31G* and (4) is the CAM-B3LYP/STO-3G*. The values marked as n/a have not been calculated owing to their excessive computer requirements

Length	(1)	(2)	(3)	(4)
2	0.000	0.062	0.001	0.060
3	0.018	0.064	0.002	0.060
4	0.014	0.062	-0.012	0.052
6	-0.010	0.044	-0.022	0.045
8	-0.011	0.037	-0.028	0.036
12	n/a	0.039	n/a	0.042
16	n/a	0.042	n/a	n/a

factors were obtained from the matching of the theoretical and experimental frequencies. Note that the band intensities do not play any role for the calculation of the polaron binding energy (see Eq. 2). For the purposes of the evaluation of the electron-phonon term (see Eq. 2), both spectra (for the *neutral* and *relaxed* states) were scaled with the same scaling factor which was determined from the data in Table 2.

Polaron binding energy

Tables 3 and 4 contain the calculated values of both terms in Eq. 1. It is clear that the contribution of the deformation energy (E_{def}) to the polaron binding energy (E_p) is about ten-fold greater than the contribution of the electron-phonon term ($E_{e\text{-}ph}$). This contradicts the results presented for molecular crystals [6]. The difference arises from the fact that polymer chains are usually softer and are more deformed than a crystal lattice under the influence of the excess charge. The absolute values of the deformation energy are in good agreement for both basis sets for chains longer than three monomer units. On the other hand, the electron-phonon term calculated with the STO-3G* basis

set is about three times greater than the one obtained using the 6-31G* basis set, which even yields negative values for 6- and 8-mer. Since the $E_{e\text{-}ph}$ term is given by the $\omega_0^{(k)} - \omega_p^{(k)}$ difference, the negative outcome would suggest an increase in the $\omega_p^{(k)}$ value, which could be interpreted as bond strengthening. In fact, some of the bonds in the polymer backbone are shortened upon electron removal. Since the infrared spectrum with the 6-31G* basis set could be calculated for only a few different lengths, we cannot determine if this effect is a fluctuation, an error in the calculations or consistent behavior. In general, both terms tend to decrease with increasing chain length.

The dependence of the deformation and polaron binding energy on the chain length is shown in Fig. 7. The values for both of the basis sets utilized exhibit a steadily decreasing dependency for chain lengths greater than three monomer units. The results of the B3LYP method are significantly lower than the CAM-B3LYP values. The reason for this is the fact that the former method over-estimates the distribution of the polaron on the chain [20]. The first two points for both basis sets are obviously out of range because of the extreme deformations that take place on such short chains. From the qualitative point of view, both basis sets provide similar results.

As follows from Fig. 7, the value of the polaron binding energy is strongly dependent on chain length, i.e., the delocalization length, which has been studied by spectroscopic methods in poly(methyl-n-propylsilane) [22]; the value of 16 monomer units was found to be the most probable extent of the delocalization. However, this number is only a rough estimate and the actual value can vary within a range of several monomer units (e.g., molecular dynamics simulations for PMPSi have resulted in an approximate length of 20 monomer units [2]). Since the values of the electron-phonon term for such long chains could not be calculated for the 6-31G* basis set, we have taken the value of $E_{e\text{-}ph}$ for the 8-mer as an approximate estimation. The results for the STO-3G* basis set (see Table 4) suggest that this term does not change significantly

Fig. 7 (a) The dependence of the deformation energy (E_{def}) on the chain length (the number of Si atoms in the backbone, N); (b) the dependence of the polaron binding energy (E_p) on the chain length (the dotted lines are an estimate, where the unknown electron-phonon term was substituted by the value for 8-mer). The results have been obtained at the B3LYP/6-31G* (■), B3LYP/STO-3G* (▲), CAM-B3LYP/6-31G* (□) and CAM-B3LYP/STO-3G* (△) levels
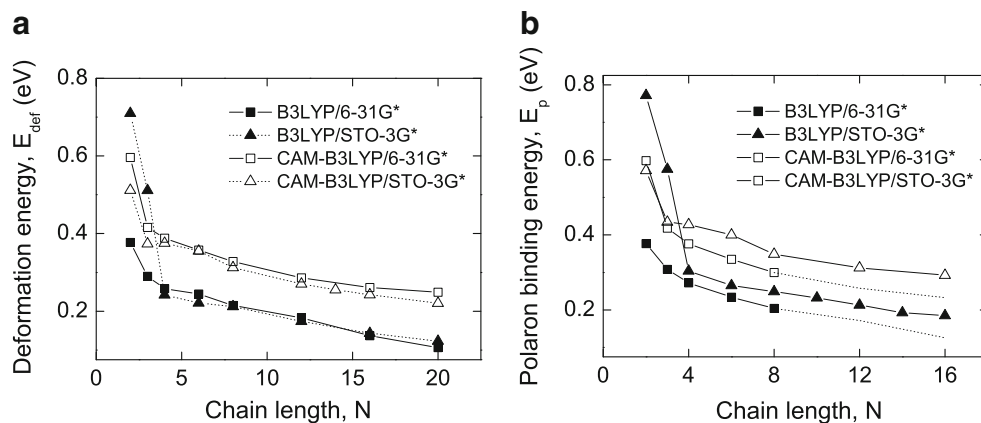
in the range of lengths eight through 16 monomer units. In accord with this fact, the values of the polaron binding energy obtained was 0.13 eV and 0.23 eV at the B3LYP/6-31G* and CAM-B3LYP/6-31G* levels of theory, respectively. The difference between the values can be explained by the fact that the CAM-B3LYP method provides a better model for long-range interactions, which plays a significant role in polaron calculations, since it extends over several monomer units. We therefore conclude that the result of this method is more reliable.

Conclusions

Hybrid DFT (B3LYP and CAM-B3LYP methods) calculations were used to estimate the polaron binding energy in the poly[methyl(phenyl)silylene] polymer. The potential well causes the charge carrier's self-trapping and the decrease of the charge-carrier mobility. There is a decreasing dependence of the polaron binding energy on the polymer chain length. The analysis shows that for the most probable delocalization length of 16 monomer units, which was obtained experimentally, the polaron binding energy is about 0.23 eV (calculated at the CAM-B3LYP/6-31G* level of theory). This result falls in the middle of the range of values obtained by experiments.

References

1. Miller RD, Michl J (1989) Polysilane high polymers. Chem Rev 89:1359–1410
2. Němec H, Kratochvílová I, Šebera J, Kochalska A, Nožár J, Nešpůrek S (2010) Charge carrier mobility in poly[methyl(phenyl) silylene] studied by time-resolved terahertz spectroscopy and density functional theory. Phys Chem Chem Phys 7:2850-2856
3. Nešpůrek S, Valerian H, Eckhardt A, Herden V, Schnabel W (2001) Charge carrier transport in poly[methyl(phenyl)silylene]: the effect of additives. Polym Adv Technol 12:306–318
4. Perry RJ, Bässler H, Borsenberger PM (1994) Charge transport in poly(methylphenylsilane): the case for superimposed disorder and polaron effects. J Polym Sci B Polym Phys 32:1677–1685
5. Pan L, Zhang M, Nakayama Y (1999) Effect of residual solvent on carrier transport in polysilane. J Chem Phys 110:10509–10514
6. Capek V, Silinsh EA (1994) Organic Molecular Crystals: Interaction, Localization, and Transport Phenomena. AIP Press, NY
7. Jang JW, Lee CE, Toman P, Nešpůrek S (2002) Conformation changes of polysilanes during the polaron formation. Curr Appl Phys 2:327–330
8. Sworakowski J, Nespurek S, Toman P (2003) Charge carrier transport on molecular wire controlled by dipolar species: towards light-driven molecular switch. Thin Solid Films 438:268–278
9. Becke AD (1992) Density-functional thermochemistry. I. the effect of the exchange-only gradient correction. J Chem Phys 96:2155–2161
10. Becke AD (1992) Density-functional thermochemistry. II. the effect of the perdew-wang generalized-gradient correlation correction. J Chem Phys 97:9173–9178
11. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 98:5648–5653
12. Handy NC, Yanai T, Tew DP (2004) A new hybrid exchange-correlation functional using the coulomb-attenuating method (cam-b3lyp). Chem Phys Lett 393:51–57
13. Silinsh EA (1980) Organic Molecular Crystals: Their Electronic States. Springer, Berlin
14. Frisch MJ et al (2009) Gaussian 09. Gaussian Inc, Pittsburgh, PA
15. Pople JA, Redfern PC, Rassolov VA, Ratner MA, Curtiss LA (2001) 6-31 g* basis set for third-row atoms. J Comput Chem 22:976–984
16. Hehre WJ, Stewart RF, Pople JA (1969) Self-consistent molecular-orbital methods. I. use of gaussian expansions of slater-type atomic orbitals. J Chem Phys 51:2657–2664
17. Takeda K, Teramae H (1989) Ab initio studies on silicon compounds. Part II. The gauche structure of the parent polysilane. J Am Chem Soc 111:1281–1285
18. McCrary VR, Sette F, Chen CT, Lovinger AJ, Robin MB, Stöhr J, Zeigler JM (1988) Polarization effects in the valence and inner-shell spectra of poly(di-n-hexylsilane). J Chem Phys 88:5925–5933
19. Toman P (2000) Quasiparticles in σ-conjugated polymers. Synth Met 109:259–261
20. Zhang IY, Wu J, Xu X (2010) Extending the reliability and applicability of b3lyp. Chem Commun 46:3057–3070
21. Merrick JP, Moran D, Radom L (2007) An evaluation of harmonic vibrational frequency scale factors. J Phys Chem A 111:11683–11700
22. Irie M, Irie S (1997) Absorption spectra of radical ions of low molecular weight poly(methyl-n-propylsilane)s chain length dependence. Macromolecules 30:7906–7909

Binding of BIS like and other ligands with the GSK-3β kinase: a combined docking and MM-PBSA study

Nihar R. Jena

Abstract Binding of several bisindolylmaleimide (BIS) like (BIS-3, BIS-8 and UCN1) and other ligands (H89, SB203580 and Y27632) with the glycogen synthase kinase-3 (GSK-3β) has been studied using combined docking, molecular dynamics and Poisson-Boltzmann surface area analysis approaches. The study generated novel binding modes of these ligands that can rationalize why some ligands inhibit GSK-3β while others do not. The relative binding free energies associated with these binding modes are in agreement with the corresponding measured specificities. This study further provides useful insight regarding possible existence of multiple conformations of some ligands like H89 and BIS-8. It is also found that binding modes of BIS-3, BIS-8 and UCN1 with GSK-3β and PDK1 kinases are similar. These new insights are expected to be useful for future rational design of novel, more potent GSK-3β-specific inhibitors as promising therapeutics.

Keywords BIS like inhibitors · Docking · Glycogen synthase kinase-3β (GSK-3β) · MM-PBSA · Protein-drug binding

Introduction

Glycogen synthase kinase-3 (GSK-3) is a serine-threonine kinase [1, 2] that plays various important physiological roles such as signal transduction, cell proliferation, survival and differentiation, apoptosis, transcription, insulin action etc. [3–9] in several organisms. However, presumably, over expression of the GSK-3 kinase activity is lethal [10], which has been implicated in different pathological conditions like cancer, neurological disorders, diabetes, and bipolar disorders etc. [11–13]. Association of GSK-3 with various diseases has made it an attractive potential therapeutic target in recent days [1, 14, 15], which have paramount importance in medicinal chemistry [15, 16].

Recently binding affinities of several ligands, e.g., H89, SB203580, Y27632, BIS-3, BIS-8 and UCN1 (Scheme 1) complexed with GSK-3β have been studied by Davies et al. [17]. These ligands are mainly adenosine triphosphate (ATP) competitors [1, 15, 16, 18–20]. It has been found that among these ligands, BIS-8 is associated with highest binding affinity [20]. However, the reason of this specificity has not yet been understood clearly. Furthermore, it is believed that H89, SB203580, Y27632 and UCN1 are specific inhibitors of PKA [21, 22], p38 [23–26], ROCK [27–43] and PDK1 [30] kinases respectively and are helpful for treatment of cancer, Alzheimer and other diseases. Given the fact that all kinases have similar sequence identities at the binding site, it is not understood why above ligands do not inhibit GSK-3β strongly.

Although Page and Bate [31] have tried to simulate binding affinities of above ligands complexed with GSK-3β by using molecular dynamics (MD) and Poisson-Boltzmann surface area analysis (MM-PBSA) studies [32–35], they have failed to reproduce experimentally measured specificities of these ligands [20]. This discrepancy between the experimental and theoretical studies might have arisen due to the starting structure used for the study that was generated by superposing a ligand on another ligand whose complex structure with GSK-3β is experimentally known.

N. R. Jena (✉)
Molecular Biophysics Unit, Indian Institute of Science,
Bangalore 560 012, India
e-mail: nrjena@mbu.iisc.ernet.in
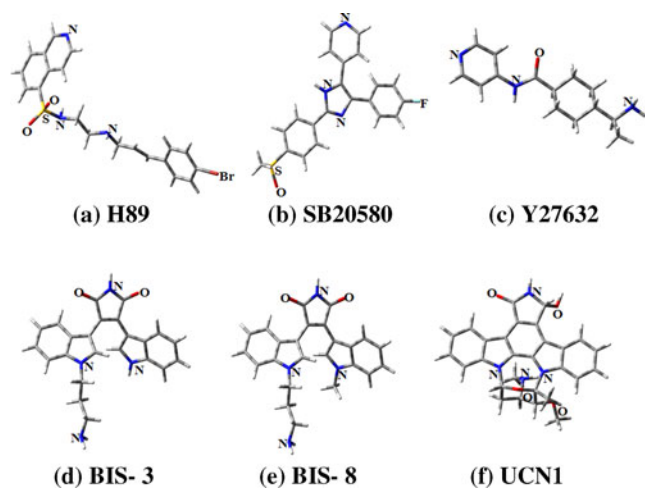
Scheme 1 Structures of different ligands considered in the present study

This method of creating initial structure is very crude and may not yield proper results always. For example, in case of structurally divergent ligands, the superposition method may lead to confusing results. Furthermore, it is found that several inhibitors bind to the kinase in multiple conformations [36, 37], which cannot be obtained by the superposition method. However, it is now increasingly believed that, where experimental kinase-inhibitor structures are not known, combined docking, MD and MM-PBSA studies can produce inhibition activities with greater and accurate precision [38–40].

Owing to the above facts, binding of H89, SB203580, Y27632, BIS-3, BIS-8 and UCN1 ligands with the GSK-3β kinase has been studied here by using combined docking, MD and MM-PBSA approaches. Two major goals of the present study are (1) to characterize whether combination of both docking, MD and MM-PBSA methods can reproduce experimentally measured specificities of above ligands for GSK-3β and (2) to suggest accurate binding modes of these ligands with GSK-3β for which experimental three dimensional structures are not available. The study is also expected to answer why some ligands can inhibit other kinases but not GSK-3β and to identify important residues that may control GSK-3β specificity.

Computational methodology

Input preparation and docking

It has been established that the ability of a good docking program is to reproduce the experimental binding mode of a complex structure with low (usually <2.0 Å) root mean square deviation (RMSD) [41, 42]. In order to validate the

docking protocol used in the present study, STU was docked first into the binding site of GSK-3β to ensure that the experimentally observed binding mode of GSK-3β-STU complex [17] is reproduced. To do so, atomic coordinates of GSK-3β and STU were obtained from the X-ray crystal structure of the GSK-3β complex (resolution 2.20 Å) deposited in the Protein Data Bank (PDB) (PDB code 1q3d) [17]. Hydrogen atoms to the kinase and ligand were added by using the all-atom Amber (version 9) [43] and Discovery studio (DS Viewer Pro 6.0) molecular modeling program respectively. All water molecules that were present in the X-ray crystal structure of the above complex were allowed to spin in the docking calculation by turning on the toggle option of the GOLD (version 5.0) Program [41, 42, 44–47]. The binding site of the GSK-3β kinase was defined to include all kinase residues located within a sphere of radius 10 Å from the center of the ligand. This selection of the binding site was aimed to provide adequate space for proper conformational sampling of the ligand. The number of operations was set to 2×10^6 with a total of GA 100 runs. All other docking parameters were set to the default values of the GOLD program. It should be mentioned that, although during docking calculation, the receptor (GSK-3β) was held rigid; the dihedrals of the protein residues containing OH and NH_3^+ groups were optimized [41, 42]. However, the ligand was treated flexible and allowed to move in all possible dimensions. Further, during docking, all torsion angles of the ligand and ring geometries were optimized [41, 42]. This helped STU to take neutral, non-biased conformation within the kinase binding site.

Use of above docking protocol was found to reproduce the experimental binding mode of GSK-3β-STU complex with RMSD of ~0.4 Å with respect to the experimental structure [17]. This protocol was then generalized to dock other ligands studied here into the binding pocket of GSK-3β. It should be mentioned that atomic coordinates of H89, SB203580, Y27632, BIS-3, BIS-8 and UCN1 were taken from the complex crystal structures of PKA-H89 (code 1ydt), p38-SB203580 (code 1a9u), PKA-Y27632 (code 1q8t), PDK1-BIS-3 (code 1uu9), PDK1-BIS-8 (code 1uvr) and PDK1-UCN1 (code 1okz) respectively. By repeating above procedure used for docking of STU, above ligands were docked into the binding site of GSK-3β.

Molecular dynamics (MD) simulation

In order to obtain more accurate binding modes and energies, several docked conformations of above ligands were considered for subsequent MD and MM-PBSA studies. The selected conformations were met the following conditions: (1) they had higher ranks in general (2) the adenine mimetic ring of H89, SB203580 and Y27632 was

oriented in a similar manner to that of AMP-PNP, as observed in its complex structure with GSK-3β [17], (3) the 5-membered head group of BIS-8, BIS3 and UCN1 was oriented toward the hinge loop of GSK-3β as observed in the GSK-3β-STU complex structure [17].

Each protein-ligand complex was solvated in a rectangular box of TIP3P water molecules [48] with a minimum solute wall distance of 10 Å after neutralizing the system by adding counterions. The electrostatic was treated by applying periodic boundary condition with the particle-mesh Ewald method [49]. The sander module of Amber 9 [47] was used for the simulation, where a residue-based cut off of 10 Å, a time step of 2 fs and a constraint of bond lengths involving hydrogen atoms using the SHAKE algorithm [50] were employed. The nonbonded pair list was updated in every 50 fs. The bcc [51] partial atomic charges and generalized amber force field parameters (GAFF) [52, 53] were used for the ligands by using the antechamber module of Amber 9 [47] program. The ff03 [54] force field parameters were assigned to the kinase under investigation.

The solvated complexes were minimized with 1000 steps of steepest descent minimization followed by 1000 steps of conjugate gradient minimization. Minimization of the complex was carried out by restraining the protein backbone (CA, C, O, H) with force constants of 100 and 20 kcal mol^{-1} Å^{-2} in two steps. In the subsequent step, the complex was heated slowly from low temperature to high temperature i.e. 300 K, keeping the force constant of 20 kcal mol^{-1} Å^{-2} intact on the protein backbone at a constant volume (NVT) for 20 ps. The whole system was then equilibrated at 300 K for 1 ns keeping the force constant of 20 kcal mol^{-1} Å^{-2} intact on the protein backbone. However, the kinase side chains were kept flexible and allowed to take neutral representations during dynamics. The main aim of the restraint MD was to keep the protein structure as close as to the experimental structure. This type of restraint MD has been suggested to yield encouraging results [55, 56]. The 1 ns simulation resulted stable trajectories as was evident from the RMSD calculation of the protein structure (< 1 Å) with respect to the minimized structure in all cases studied here.

Relative binding free energy calculation by MM-PBSA approach

The binding free energies were calculated by using the MM-PBSA free energy approach [32–35]. For this purpose, last 20 snapshots were taken at 5-ps intervals from the final trajectory, and interaction energies were calculated using the scripts provided with the Amber suite [47]. The binding free energy of the kinase-ligand binding (ΔG$_{bind}$) was determined by taking the difference between the free energies of the kinase-ligand complex (G$_{comp}$) and the unbound kinase (G$_{kinase}$) and ligand (G$_{ligand}$) as follows:

$$\Delta G_{bind} = G_{comp} - G_{kinase} - G_{ligand}. \tag{1}$$

The binding free energy (ΔG$_{bind}$) contains contributions from sum of the changes in the molecular mechanical (MM) gas-phase binding energy (ΔE$_{MM}$), entropic contributions (-TΔS) and solvation free energy (ΔG$_{sol}$):

$$\Delta G_{bind} = \Delta E_{MM} - T\Delta S + \Delta G_{sol}, \tag{2}$$

where

$$\Delta E_{MM} = \Delta E_{inter} + \Delta E_{elc} + \Delta E_{vdw}, \tag{3}$$

and

$$\Delta G_{sol} = \Delta G_{polar} + \Delta G_{nonpolar}. \tag{4}$$

The ΔE$_{MM}$ that contains sum of the contributions from internal (ΔE$_{inter}$), electrostatic (ΔE$_{ele}$) and van der Waals energies (ΔE$_{vdw}$) was calculated by using the Amber 9 program [24]. The polar contribution (ΔG$_{polar}$) to the solvation free energy was calculated by using the Poisson-Boltzmann (PB) equation, where dielectric constants of 1 and 80 were used for the inner and outer sides of the solute cavity. The non-polar contribution (ΔG$_{nonpolar}$) to the solvation free energy was calculated from the solvent-accessible surface area (SASA), i.e.,

$$\Delta G_{polar} = \Delta G_{PB} \text{ and } \Delta G_{non-polar} = \gamma SASA + \beta. \tag{5}$$

The SASA was determined by using the Molsurf program [57], where the surface tension proportionality constant (γ) and the free energy of non-polar solvation for a point solute (β) were taken to be 0.00542 kcal mol^{-1} Å^{-2} and 0.92 kcal mol^{-1} respectively. All other parameters were considered to be default parameters as defined in Amber 9 [47]. As I was mainly concerned about binding of similar ligands to one particular receptor (GSK-3β), entropic contributions to the free energy (TS) were not evaluated. Due to this reason, the binding free energy will be termed as the relative binding free energy (ΔG$_{bind}^{r}$) and was estimated as follows:

$$\Delta G_{bind}^{r} = \Delta E_{MM} + \Delta G_{sol}. \tag{6}$$

It is reasonable to omit entropy from binding free energy calculations as it does not contribute much to the relative binding free energies of the similar ligands that bind to the same receptor [58–60]. Furthermore, there is no straightforward way to quantitatively [60–62] estimate the entropic contributions to binding, which are again not free from limitations in the approximations used for its computation [58, 61]. It has been further established that

without considering entropic contributions to the free energy, significantly good correlation between the calculated and measured binding affinities can be obtained [63, 64].

Results and discussion

Structure and relative binding free energy of GSK-3β-H89 complex

By applying (1) and (2) criteria as discussed above, three docked conformations of H89 (Scheme 1a) (ranks-1,6,9) were considered for subsequent MD and MMPBSA studies. These conformations of H89 are shown in Fig. S1 (Supporting information). Superposition of these conformations on each other shows that H89 can adopt different orientations with respect to the hinge loop. In rank-1, the adenine mimetic group (isoquinoline) of H89 is orientated toward the glycine-rich (G-) loop of GSK-3β while it is pointed toward the hinge loop in ranks-6 and 9. Although in rank-6, the isoquinoline group of H89 is orientated toward the hinge loop (H-loop) of the GSK-3β, it has appreciably moved away from it (Fig. S1).

The relative binding free energies of all kinase-ligand complexes obtained by the subsequent MD and MM-PBSA studies are presented in Table 1. For the sake of comparison, the corresponding free energy values obtained by Page and Bates [31] and the docking scores are also given in Table 1. It is quite evident from Table 1 that conformation of H89 corresponding to rank-1 is associated

with relatively more binding free energy than that of ranks-6 and 9. It should be mentioned that this conformation of H89 (rank-1) has not been observed experimentally for any kinase-H89 complex and hence interesting to find its existence experimentally. In this conformation, the sulfonamide group of H89 makes two direct hydrogen bonds with the Ser66 and Asp200 residues of the G- and activation (A-) loops respectively (Fig. 1). The long chain ethylamino group of H89 also makes an additional hydrogen bond with the Asp200 residue (Fig. 1). The percentage occupations of these hydrogen bonds are more than 80% for the last 100 ps equilibration. Moreover, it is also found that the Br atom of the H89 can occasionally make a hydrogen bond with the Val135 hinge residue. Besides these hydrogen bonding interactions, the isoquinoline group of H89 makes one stacking interaction with the Phe67 residue of the G-loop and few van der Waals interactions with other residues of the same loop.

Interestingly, the isoquinoline group of H89 faces toward the H-loop of the GSK-3β in ranks-6 and 9. This orientation of H89 has been observed in X-ray crystal structure of the PKA-H89 complex [21]. Further, it is found that the relative orientations of isoquinoline and sulfonamide groups of H89 in ranks-6 and 9 are similar; the only difference arises mainly due to dissimilar orientations of the two NH groups of the long chain of H89. In rank-6, these two NH groups are pointing toward the Asp200 residue similar to that found in rank-1 (Fig. 1). However, in rank-9, only one NH group is pointing toward the Asp200 residue. As a result, H89 in rank-6 can make two hydrogen bonds with the Asp200 residue, while in

Table 1 Relative binding free energies (ΔG_{bind}^r) (kcal mol^{-1}) involved in the binding of GSK-3β with the H89, SB203580, Y27632, BIS-3, BIS-8, and UCN1 ligands obtained at room temperature (298.13 K) and one atmospheric pressure (1 atm) when the protein backbone (CA, C, O, H) is restrained with k=20 kcal mol^{-1} Å^{-2}

Ligands	Rank	Docking score	ΔE_{MM}	ΔG_{sol}	ΔG_{bind}^r	$\Delta G_{bind}^r{}^a$
H89	1	68.14	−69.94	45.33	−24.62	−36.61
	6	63.78	−60.66	38.91	−21.75	
	9	63.50	−59.55	42.00	−17.56	
SB203580	1	69.06	−38.86	18.72	−20.14	−26.21
	37	63.99	−37.14	22.43	−14.71	
	82	59.69	−45.62	21.74	−23.88	
Y27632	1	54.34	−36.71	26.75	−9.96	−21.51
	11	51.87	−46.48	26.46	−20.02	
BIS-3	1	69.38	−61.49	36.27	−25.22	−39.04
	4	68.82	−72.32	37.02	−35.31	
	10	66.19	−65.68	39.56	−26.13	
BIS-8	1	77.02	−63.89	30.37	−32.12	−29.38
	3	76.21	−69.69	34.08	−35.62	
	4	75.55	−73.64	42.02	−33.52	
	5	75.17	−74.88	32.95	−41.93	
	9	74.66	−72.59	32.88	−39.72	
	51	68.97	−67.45	31.32	−36.13	
UCN1	1	34.94	−70.12	34.14	−35.98	−26.98

^{a} From [31]

Fig. 1 MD-simulated structure of the most stable GSK-3β-H89 (rank-1) complex. The GSK-3β is shown in the cartoon representation and hydrogen bonding interactions between the kinase and ligand are shown in dotted lines. The H-loop, G-loop, C-loop and A-loop correspond to hinge, glycine-rich, catalytic and activation loops of the kinase respectively
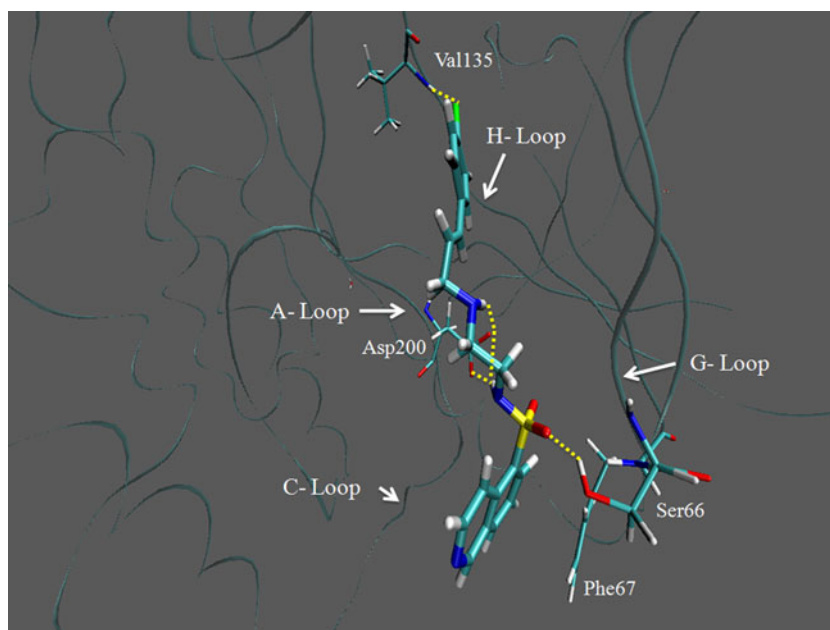
rank-9, it can make only one hydrogen bond with the same residue. It is worth mentioning that the isoquinoline group of H89 in rank-6 that was shifted away from the hinge loop of GSK-3β after docking computation moves toward the hinge loop during MD study, as a result, the N-H (Val135; GSK-3β)–N (isoquinoline; H89) distance is same in both ranks-6 and 9. Nevertheless, the isoquinoline group of H89 in these conformations failed to hydrogen bond with the GSK-3β kinase unlike in PKA-H89 [21] and GSK-3β-AMP-PNP [17] complexes. Moreover, the bromophenyl group of H89, which is supposed to mimic the triphosphate group of ATP, also fails to luck the

catalytic loop (C-loop) of GSK-3β. However, in this conformation, the bromophenyl group turns toward the G-loop and makes few van der Waals interactions with the G-loop residues. In this conformation, H89 failed to make any stacking interaction with the Phe67 residue of the G-loop as obtained in rank-1.

It should be mentioned that though H89 in rank-6 binds with GSK-3β in a similar fashion as it binds with PKA (pdb 1ydt) [21], due to lack of two key hydrogen bonding interactions equivalent to H89-Val123 and H89-Glu170 (Fig. 2) of PKA, H89 cannot inhibit GSK-3β as strongly as PKA. Besides these key hydrogen bonding interactions,

Fig. 2 **a** Superposition of GSK-3β-H89 (rank-6) (*in yellow*) and PKA-H89 (pdb 1ydt) (*in cyan*) complexes. For comparison, binding modes of H89 (*in red* and *yellow*) in the active sites of **b** PKA and **c** GSK-3β are also shown. Direct hydrogen bonding interactions between the kinase and ligand are shown in dotted lines
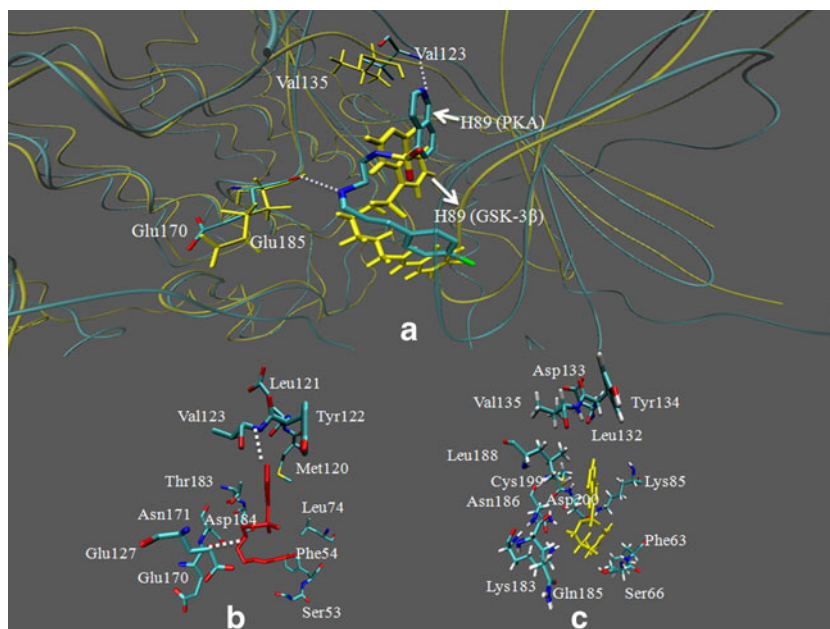
H89 in GSK-3β also cannot interact with C- and G-loop residues (Fig. 2b,c). However, as conformation of H89 in rank-1 is associated with higher relative binding free energy and in rank-6 its binding mode with GSK-3β is similar to that of PKA; multiple conformations of H89 in the GSK-3β active site might be possible.

Structure and relative binding free energy of GSK-3β-SB203580 complex

Three docked conformations of SB203580 (Scheme 1b) (ranks-1,37,82) considered for subsequent MD and MM-PBSA studies by applying (1) and (2) criteria as discussed earlier are displayed in Fig. S2 (Supporting information). Superposition of these structures on each other suggests that, while the two head groups attached to the imidazole ring are freely rotable and can take different orientations with respect to the hinge loop, the methyl sulfonyl group always extends toward the G-loop. The relative binding free energies of these complexes presented in Table 1 indicate that conformation of SB203580 in rank-82 has relatively more binding free energy than the other two conformations. This is due to the fact that in this conformation, the pyridine group of SB203580 can make one hydrogen bonding interaction with the Val135 H-loop residue (occupancy 80%) (Fig. 3). In addition, the fluorophenyl and long chain methyl sulfonyl groups can also make few hydrophobic interactions with the H- and C-loop residues respectively (Fig. 3b). However, in other conformations, SB203580 failed to interact with the kinase completely. It should also be mentioned that although in rank-82, the SB203580-Val135 hydrogen bonding interaction is equivalent to the

corresponding interaction involved in the GSK-3β-AMP-PNP complex, the methyl sulfonyl group of SB203580 failed to mimic other two hydrogen bonding interactions between the phosphate groups of AMP-PNP and GSK-3β [17].

If we compare binding modes of SB203580 in GSK-3β-SB204580 and p38-SB203580 complexes, it is clear that in the former complex, although the imidazole and fluorophenyl rings of SB203580 face toward the β- and H-loop regions of the kinase respectively, they cannot interact with Lys58 and Thr138 residues of those regions respectively. However, the corresponding interactions of SB203580 in the later complex (i.e. imidazole-Lys53 and fluorophenyl-Thr106) have been found to be crucial for inhibitory activities of p38 [23] and c-Raf kinases [65]. Further, in the former complex, the cyclopropylmethyl group of SB203580 failed to mimic necessary hydrophobic interactions with the G-loop residues of GSK-3β unlike as observed in the later complex [23]. Although the pyridine-Val135 hydrogen bonding interaction in the GSK-3β-SB204580 complex is similar to the corresponding interaction observed in the p38-SB203580 complex [23], it is not enough to inhibit GSK-3β strongly in agreement with the experimental finding [20, 23, 65].

Structure and relative binding free energy of GSK-3β-Y27632 complex

The two docked conformations of Y27632 (Scheme 1c) (ranks-1,11) considered for subsequent MD and MM-PBSA studies by applying (1) and (2) criteria as mentioned above are shown in Fig. S3 (Supporting information). In ranks-1

Fig. 3 MD-simulated structure of the most stable GSK-3β-SB203580 (rank-82) complex. (a) Orientation of SB203580 in the binding site of GSK-3β involving the hinge (H-loop), glycine-rich (G-loop), catalytic (C-loop) and activation (A-Loop) loops. (b) Binding of SB203580 (in red) with the key residues of GSK-3β. The hydrogen bond between the ligand and kinase has been depicted in dotted line
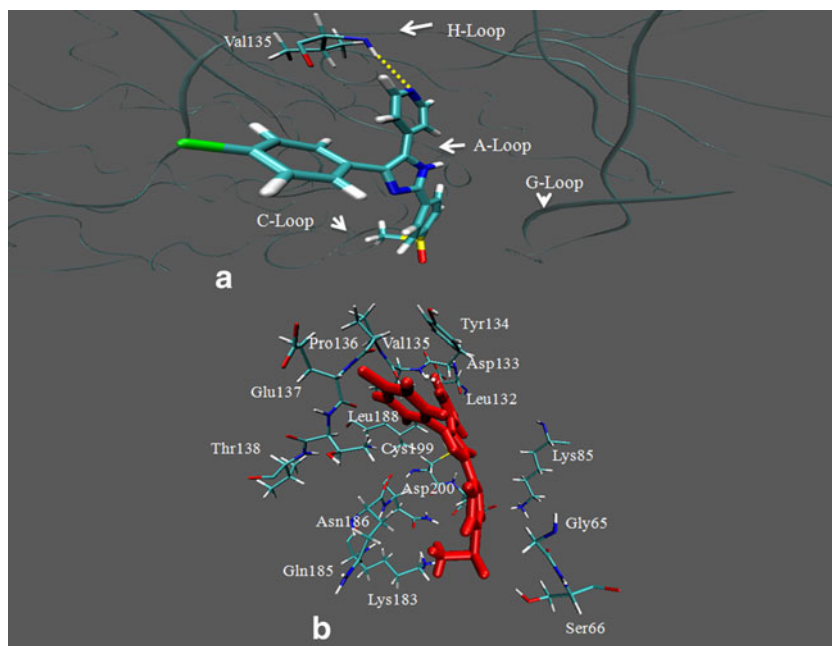
Fig. 4 MD-simulated structure of the most stable GSK-3β-Y27632 (rank-11) complex. (**a**) Orientation of Y27632 in the binding site of GSK-3β involving the hinge (H-loop), glycine-rich (G-loop), catalytic (C-loop) and activation (A-Loop) loops. (**b**) Binding of Y27632 (in blue) with the key residues of GSK-3β. The hydrogen bonding interaction between the ligand and kinase has been depicted in dotted line
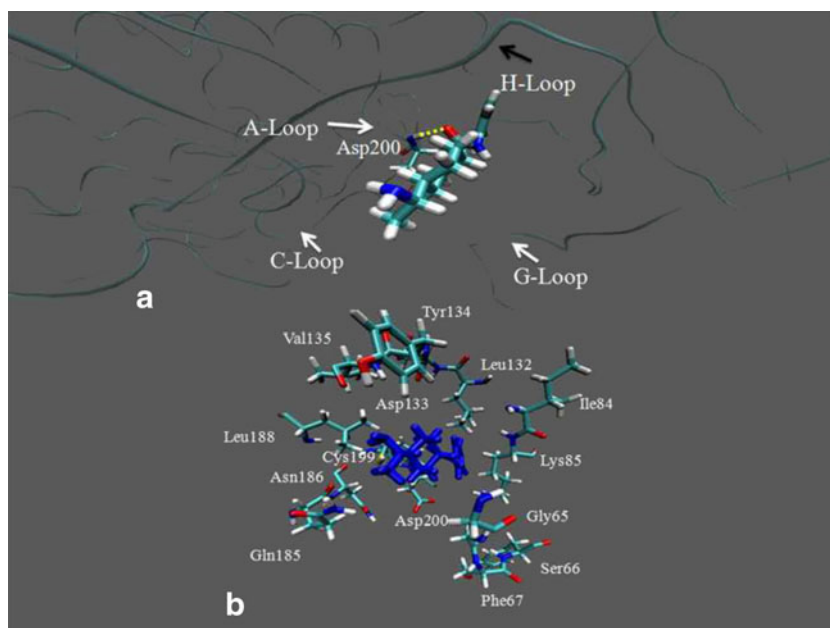

and 11 the pyridine ring of Y27632 points toward the H- and G-loops respectively. According to the MM-PBSA study, the GSK-3β-Y27632 complex corresponding to rank-11 is almost twice more stable than the same complex corresponding to rank-1 (Table 1). This is due to the fact that in the former complex, Y27632 makes one strong hydrogen bond with the Asp200 residue (occupancy 100%) of the A-loop (Fig. 4), while in other complex, Y27632 does not make any contact with the kinase. It should be mentioned that although, Y27632 in rank-82 makes one direct contact with one of the A- loop residues, it has been pulled away from the H-, C- and G-loop regions (Fig. 4). Nevertheless, the pyridine ring of Y27632 can make hydrophobic interaction with the Leu132 hinge residue (Fig. 4b). However, in the GSK-3β-AMP-PNP complex, the adenine and phosphate groups of AMP-PNP have been observed to make two direct hydrogen bonds each with the Asp133 and Val135 hinge residues and Gln185 (G-loop) and Lys85 (β-strand) respectively [17].

If we compare binding modes of Y27632 in GSK-3β-Y27632 and PKA-Y27632 complexes, it is clear that in the former complex, Y27632 does not make any hydrogen bonding contact with the GSK-3β kinase. However, in the later complex the pyridine and long chain amide groups make two hydrogen bonds with the Val123 and Thr51 or Asn171 residues of PKA [29] respectively. In addition, as Y27632 has moved away from different binding regions of GSK-3β, it cannot also mimic other polar and hydrophobic interactions corresponding to PKA-Y27632 complex. This indicates that binding of Y27632 with GSK-3β is not as strong as its binding with PKA. This may be the reason, why Y27632 cannot inhibit GSK-3β kinase.

Structures and relative binding free energies of GSK-3β-BIS-3 and GSK-3β-BIS-8 complexes

Structures of the two bisindolylmaleimide (BIS) analogues i.e. BIS-3 (Scheme 1d) and BIS-8 (Scheme 1e) are similar to that of STU. The main difference between BIS analogues and STU arises due to the flexibility of the two indole rings present in the BIS analogues that enable them to freely rotate in the kinase binding pocket to make suitable interactions to stabilize the resulting complex. Structures of above two BIS analogues differ from each other minutely, i.e. the indole N-H group in BIS-3 has been replaced by the indole N-CH$_3$ group in BIS-8 (Scheme 1d,e). This minor modification significantly affects their kinase inhibitory activities (IC$_{50}$ values of BIS-3 and BIS-8 for PDK1 are 1_μM and 4_μM respectively) [30]. Though these inhibitors are specific to PKC kinase, they also inhibit other kinases like PDK1 [30], GSK-3β [20] etc.

Superposition of different docked conformations of BIS-3 (ranks-1,4,10) considered for subsequent MD and MM-PBSA studies by applying (1) and (3) criteria as discussed earlier is displayed in Fig. S4 (Supporting information). In all conformations, the maleimide head group is facing toward the H- loop and the flexible indole group takes different orientations. According to the subsequent MD and MM-PBSA studies, conformation of BIS-3 in rank-4 forms the most stable complex with the GSK-3β kinase (Table 1). This can be understood as follows. The hydrogen bonding analysis of different complex structures formed between GSK-3β and BIS-3 suggest that in the complexes corresponding to rank-1 and 10, the maleimide head group moves away from the hinge loop and hence unable to interact with it. However, the α-indole and long

Fig. 5 MD-simulated structure of the most stable GSK-3β-BIS-3 (rank-4) complex. (**a**) Orientation of BIS-3 in the binding site of GSK-3β involving the hinge (H-loop), glycine-rich (G-loop), catalytic (C-loop) and activation (A-Loop) loops. (**b**) Binding of BIS-3 (*in orange*) with the key residues of GSK-3β. The hydrogen bonding interactions between the ligand and kinase have been depicted in dotted lines
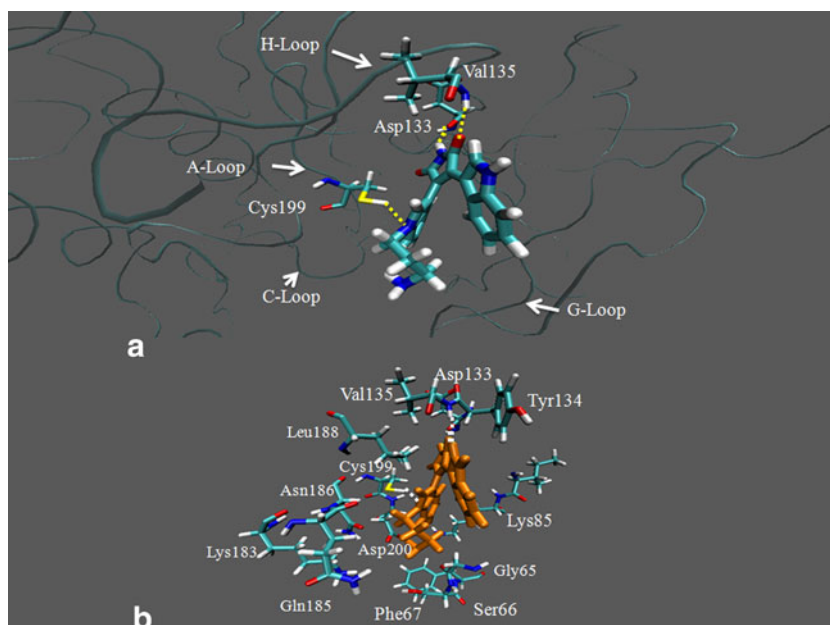
chain amide groups of BIS-3 can make direct hydrogen bonds with the Ile62 (occupancy 90%) and Lys183 residues (occupancy 90%) in rank-1, and with Ile60 (occupancy 99%) and Gly65 residues (occupancy 77%) in rank-10 respectively. In contrast to the above two complex structures, the maleimide head group of BIS-3 in rank-4 makes two strong hydrogen bonds with the Asp133 (occupancy 100%) and Val135 (occupancy 100%) residues of the H-loop (Fig. 5). These interactions are similar to the corresponding interactions of AMP-PNP and STU with the GSK-3β.

If we compare binding modes of BIS-3 in GSK-3β and PDK1, it is clear that the BIS-3-Asp133 and BIS-3-Val135 hydrogen bonding interactions of BIS-3 with GSK-3β are equivalent to the corresponding BIS-3-Ala160 and BIS-3-Ser162 interactions of the same ligand with PDK1 [30] respectively (Fig. 5 and Fig. S5; Supporting information). Due to mutation of Thr222 (PDK1) to Cys199 (GSK-3β), the maleimide head group of BIS-3 fails to make the third hydrogen bond with the Cys199 residue of GSK-3β as is observed in the PDK1-BIS-3 complex [30] (Fig. S5, Supporting information). However, this binding mode helps the N atom of the α-indole group to make an occasional hydrogen bond with the SH group of the Cys199 (occupancy 40%) (Fig. 5). Furthermore, as the terminal amide group of BIS-3 in GSK-3 extends toward the G-loop, instead of the C-loop as observed in PDK1, it failed to mimic the fourth hydrogen bonding interaction between BIS-3 and Glu166 of PDK1 [30] (Fig. S5, Supporting information). Nevertheless, binding mode of BIS-3 with the GSK-3β kinase corresponding to rank-4 is similar to its binding with PDK1 as observed in the X-ray crystal structure [30] (Fig. S5, Supporting information).

Different docked conformations of BIS-8 (ranks-1,3,4,5,9) considered for MD and MM-PBSA studies by applying criteria (1) and (3) as discussed earlier are shown in Fig. S6 (Supporting information). From this figure it is evident that in all these conformations, the maleimide head group of BIS-8 points toward the hinge loop of GSK-3β and the two indole groups take alternate orientations as found in case of BIS-3 (Fig. S4, Supporting information). Conformations of BIS-8 in ranks-1 and 5 are similar. The only difference arises due to different orientations of the terminal long chain. Similarly, conformations of BIS-8 in ranks-3 and 4 differ only in the orientation of the α-indole group that contains the methyl group. The calculated relative binding free energies as presented in Table 1 suggest that conformation of BIS-8 in rank-5 makes the most stable complex with the GSK-3β kinase.

As revealed by the MD simulation, in the case of all GSK-3β-BIS-8 complexes, the maleimide head group of BIS-8 interacts with the Asp133 (occupancy 100%) and Val135 (occupancy 100%) hinge residues of GSK-3β. These interactions are similar to the equivalent interactions observed for GSK-3β-AMP-PNP, GSK-3β-STU [17] and PDK1-BIS8 complexes [30]. In addition to these interactions, the terminal long chain amide group of BIS-8 in ranks-5 and 9 can make one additional hydrogen bond each with the Tyr134 (occupancy 56%) (Fig. 6) and Glu97 residues (occupancy 52%) (Fig. S7, Supporting information) of GSK-3β respectively. The CH$_3$ group of BIS-8 in rank-5 can further make one favorable electrostatic interaction with the Asn186 residue of the GSK-3β kinase (Fig. 6). Most interestingly, in this conformation (rank-5), the terminal long chain of BIS-8, which is supposed to mimic the phosphate chain of AMP-PNP, does not orient

Fig. 6 MD-simulated structure of the most stable conformation of GSK-3β-BIS-8 (rank-5) complex. (**a**) Orientation of BIS-8 in the binding site of GSK-3β involving the hinge (H-loop), glycine-rich (G-loop), catalytic (C-loop) and activation (A-Loop) loops. (**b**) Binding of BIS-8 (*in green*) with the key residues of GSK-3β. The hydrogen bonding interactions between the ligand and kinase have been depicted in dotted lines
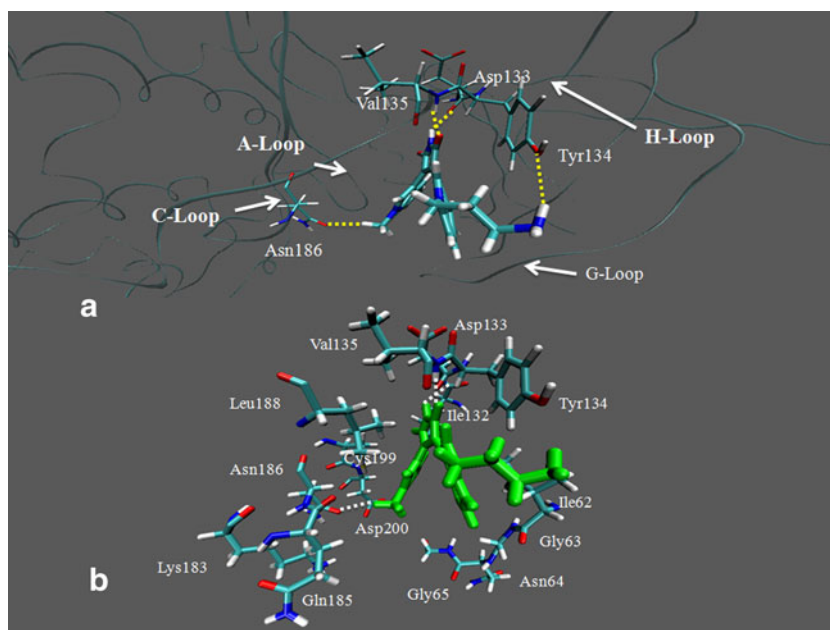
toward the C-loop [17], instead it remains within the H- and G-loops. As a result, it makes one hydrogen bonding and stacking interaction each with the Tyr134 (H-loop) and Ile62 (G-loop) residues respectively (Fig. 6). These interactions make sure that the conformation of BIS-8 in rank-5 forms the most stable complex with the GSK-3β kinase (Table 1, Fig. 6).

Although conformation of BIS-8 corresponding to rank-5 is associated with higher relative binding free energy, binding mode of BIS-8 in the GSK-3β-BIS-8 complex corresponding to rank-3 (Fig. 7a) is similar to its binding mode observed in the PDK1-BIS-8 complex (pdb 1uvr) [30] (Fig. 7b). Most interestingly, this binding mode is

similar to the binding mode of BIS-3 (rank-4) with GSK-3β (Figs. 5, 7a). In other words, the orientations of the maleimide head and two indole groups of BIS-3 (rank-4) and BIS-8 (rank-3) in the active site of GSK-3β are distinctly similar, which in turn is similar to the corresponding orientations of BIS-8 in the binding site of PDK1 (Fig. 7). The only difference arises due to the fact that the terminal long chain of BIS-8 faces toward G-loop in GSK-3β, while it points toward C-loop in PDK1 (Fig. 7). Nevertheless, this is an indication that occurrence of GSK-3β-BIS-8 (rank-3) complex cannot be ruled out and hence multiple conformations of BIS-8 in the active site of GSK-3β might be possible. This is not surprising as

Fig. 7 (**a**) MD-simulated structure of the GSK-3β-BIS-8 (rank-3) complex. (**b**) X-ray crystal structure of the PDK1-BIS-8 (pdb 1uvr) complex. The GSK-3β and PDK1 are shown in the cartoon representations and hydrogen bonding interactions between the ligand and kinases are illustrated in dotted lines. The H-loop, G-loop, A-loop and C-loop correspond to hinge, glycine-rich, activation and catalytic loops of the kinase respectively
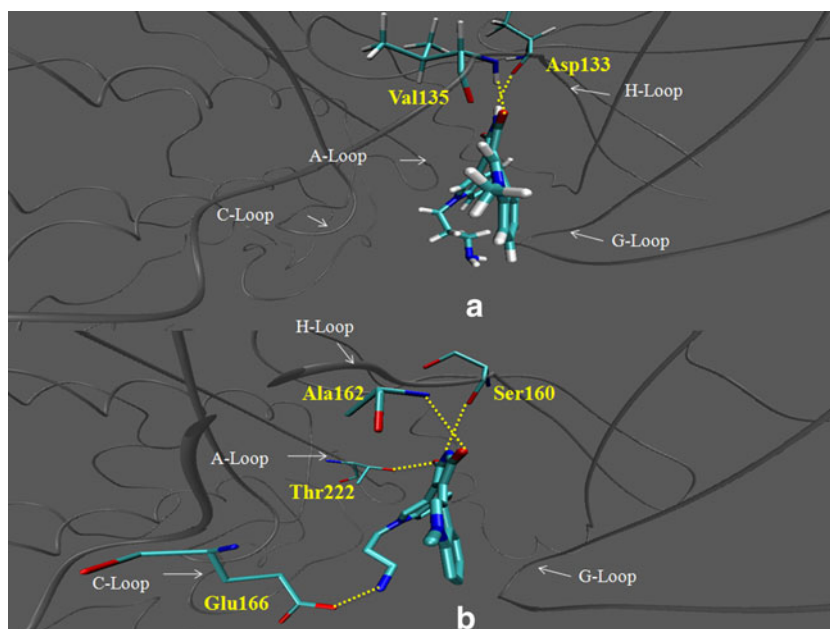
multiple conformations of various ligands have already been observed experimentally [34, 66] and modeled theoretically [67].

A comparison of the binding patterns of the most stable complexes of GSK-3β-BIS-3 (rank-4) and GSK-3β-BIS-8 (rank-5) suggest that while BIS-3 can make only three strong hydrogen bonds with the kinase, BIS-8 can make three strong hydrogen bonding interactions as well as one electrostatic and stacking interaction each with the kinase (Figs. 5, 6). This indicates that BIS-8 is more tightly bound to the GSK-3β kinase than BIS-3 in agreement with the experimental observation [20]. The extra CH_3 group in BIS-8 provides a favorable electrostatic interaction that helps BIS-8 to bind GSK-3β strongly than BIS-3 (Figs. 5, 6). However, in contrast to the experimental finding Page and Bates [31] have calculated a higher binding affinity for the GSK-3β-BIS-3 complex, which is about 10 kcal mol^{-1} more stable than the GSK-3β-BIS-8 complex (Table 1). Hence the new binding modes of BIS-3 and BIS-8 obtained in the present study are more accurate and reliable, which would play a pivotal role in understanding and designing of novel pharmacological bisindolylmaleimide derivatives specific for GSK-3β kinase [68, 69].

Structure and relative binding free energy of GSK-3β-UCN1 complex

UCN1 (Scheme 1f) is a close analogue of STU which differs from the later only at the terminal side chain. However, it differs from the above two BIS analogues in three respects, i.e., (1) the terminal side chain is different (2) the maleimide head group of BIS analogues has been

replaced by the lactam group in UCN1 (Scheme1c,d,f), (3) the indole rings of UCN1 are not freely ratable. It has been argued that these structural factors help UCN1 to acquire less entropic penalty and hence more inhibitory ability to inhibit PDK1 [30]. Although UCN1 inhibits PDK1 more strongly $\left(IC_{50} = 0.006_{\mu}M\right)$ than BIS analogues, the reverse is true for GSK-3β [20, 30, 70]. It is thus interesting to examine, if combination of docking, MD and MM-PBSA studies can explain this inhibitory activity of UCN1 properly, which has not been tackled in the previous study [31].

The docking result produced only one conformation of UCN1 (rank-1) that is very close to that of GSK-3β-STU structure [17]. This is not surprising as the experimental structure of GSK-3β-STU complex was used to dock UCN1, which shares structural similarity with STU. The subsequent MD and MM-PBSA studies suggest that the GSK3β-UCN1 complex corresponding to rank-1 is associated with ~36 kcal mol^{-1} of relative binding free energy (Table 1). From the hydrogen bonding analysis, it is clear that the lactum head group can make two strong hydrogen bonds with the Asp133 (occupancy 93%) and Val135 (occupancy 100%) hinge residues of GSK-3β (Fig. 8) like BIS-8 (Fig. 6) and STU (Fig. S8, Supporting information). These two hydrogen bonds are also equivalent to the corresponding hydrogen bonds between UCN1 and PDK1 [70] (Fig. S9, Supporting information). In addition to these interactions, the long chain amide group of UCN1 also makes one favorable hydrogen bonding interaction with the Glu185 (occupancy 88%) residue of the C-loop of GSK-3β (Fig. 8). The equivalent interaction between STU and Glu185 residue of GSK-3β mediated by a water molecule

Fig. 8 MD-simulated structure of the GSK-3β-UCN1 (rank-1) complex. (a) Orientation of UCN1 in the active site of GSK-3β involving the hinge (H-loop), glycine-rich (G-loop), catalytic (C-loop) and activation (A-Loop) loops. (b) Binding of UCN1 (in green) with the key residues of GSK-3β. The hydrogen bonding interactions between the ligand and kinase have been depicted in dotted lines
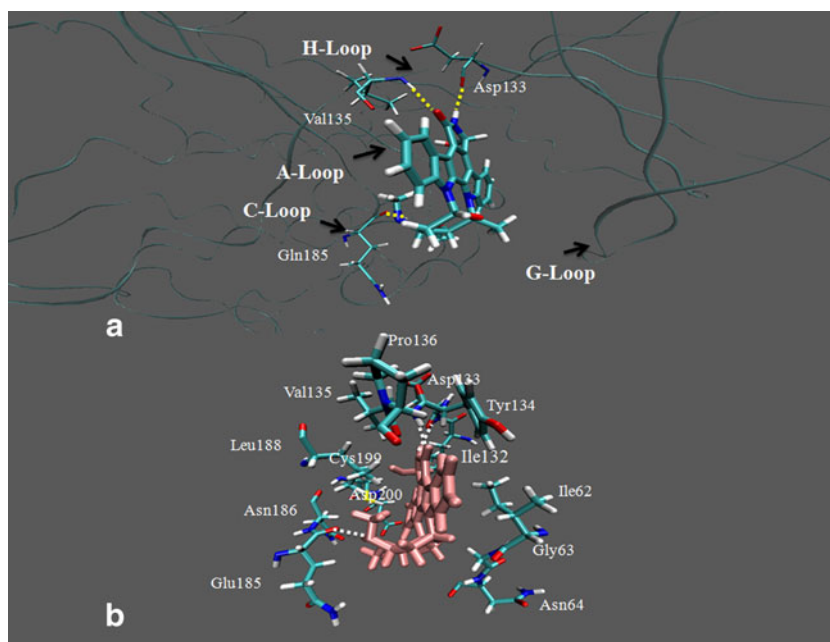
has also been observed in the GSK-3β-STU complex [17] (Fig. S8, Supporting information).

Further, the carbazole moiety of UCN1 can also make favorable electrostatic and hydrophobic interactions with the G-loop residues like GSK-3β-STU complex. This shows that STU and UCN1 bind to GSK-3β in a similar fashion. The only difference arises due to the fact that the carbonyl and N-CH_3 groups of the fused carbazole moiety of STU point up and down respectively within the cleft formed between C- and A-loops, while these groups reorient in reverse order in UCN1 (Fig. S8, Supporting information). Furthermore, if we compare binding modes of UCN1 bound to both GSK-3β and PDK1 [70], we will also find that UCN1 binds with these kinases in a similar manner (Fig. S9, Supporting information). However, those binding modes have two distinct differences. (1) The orientations of the carbonyl and N-CH_3 groups of the fused carbazole moiety of UCN1 in GSK-3β and PDK1 active sites are different as explained above. (2) The hydroxyl group of the lactum moiety of UCN1 can make a hydrogen bond with the Thr222 residue of PDK1, while it failed to interact with the Cys199 residue placed at the same position in GSK-3β (lactum-OH-Cys199=3.9 Å) (Fig. S9, Supporting information). The second difference is the same as discussed earlier in the case of BIS-8 (rank-3).

Comparison of binding modes of UCN1 and BIS8 (rank-5) bound to GSK-3β suggest that the later ligand makes three strong hydrogen bonding, one electrostatic and one stacking interactions with the GSK-3β kinase, whereas the former ligand can only make three hydrogen bonds with the GSK-3β kinase. A comparison of the hydrophobic interactions made by the BIS-8 (rank-5) and UCN1 with the GSK-3β kinase also suggests that the former is better placed in the hydrophobic pocket of GSK-3β than the later. This shows that the overall interaction of BIS-8 with GSK-3β is stronger than that of UCN1 in agreement with the experiment [20].

Specificities of different ligands

If we compare docking scores associated with all kinase-ligand complexes studied here, it is clear that BIS-8 makes the most stable complex with the GSK-3β kinase followed by BIS-3 and others (Table 1). This is further supported by the relative binding free energies of all complexes obtained by the subsequent MD and MM-PBSA studies (Table 1). As explained above, the root causes of these stabilities are interlinked with their binding modes. Further, if we compare binding of different ligands studied here with different residues of the active site, it is evident that most of the ligands such as BIS-8, BIS-3, UCN1 and SB203580 bind to GSK-3β via hinge residues (Table 2). This is in agreement with the various recent studies where it has been

Table 2 Important residues of GSK-3β that are involved in hydrogen bonding interactions with different ligands. These residues might be important for GSK-3β specificity

Inhibitors	Residues
H89	Ser66, Asp200
SB203580	Val135
Y27632	Asp200
BIS-3	Asp133, Val135, Cys199
BIS-8	Asp133, Tyr134, Val135, Asn186
UCN1	Asp133, Val135, Glu185

found that different GSK-3β specific inhibitors bind to the kinase mainly via the hinge residues [71, 72]. Nevertheless, as found in the present study, other residues like Cys199, Asp200 and Glu185 etc. (Table 2) might be important for GSK-3 specificity [73, 74]. Thus the present study not only explained specificities of H89, SB203580, Y27632, BIS-3, BIS-8 and UCN1 ligands that inhibit GSK-3β kinase, but also predicted their novel binding modes with GSK-3β. Hence combination of docking, MD and MM-PBSA studies are immensely important not only to predict accurate binding modes of kinase-ligand complexes but also to evaluate specificities of different ligands on the basis of their binding free energies.

It is worthwhile to mention that although design of ATP-like inhibitors is essential to inhibit GSK-3β activities, due to lack of specificities, these inhibitors induce un-wanted toxicities by interfering with normal functioning of other kinases. Hence exploration of GSK-3β specific inhibitors that can only target GSK-3β is very much essential. The present study is thus expected to contribute in this respect to help in understanding of important binding site residues of GSK-3β that may provide specificity to GSK-3β.

Conclusions

The combined docking, MD and MM-PBSA studies have resulted in new binding modes of H89, SB203580, Y27635, BIS-3, BIS-8 and UCN1, concerning how these ligands would bind to GSK-3β. The relative binding free energies associated with these binding modes are in agreement with the corresponding measured specificities. This shows that the proposed binding modes are accurate and reliable. A comparison of the proposed binding modes of different ligands with the corresponding experimentally observed kinase-ligand binding modes rationalizes why some ligands can inhibit GSK-3β while others cannot. It is worthwhile to mention that ligands such as BIS-8, BIS-3, UCN1, and SB203580, those bind to the hinge residues of GSK-3β, which is an important

part of the binding site, can inhibit GSK-3β, though with altered potencies. This study also provides useful insights regarding possible existence of multiple conformations of some ligands like H89 and BIS-8. It is further revealed that binding modes of BIS-3, BIS-8 and UCN1 with GSK-3β and PDK1 kinases are similar. These new insights are expected to be useful for future rational design of novel, more potent GSK-3β-specific inhibitors as promising therapeutics.

References

1. Shinsuke K, Shigeuki C, Makoto M (2008) GSK-3 inhibitors: recent developments and therapeutic potential. Curr Signal Transduction Ther 3:95–205

2. Mukai F, Ishiguro K, Sano Y, Fujita SC (2002) Alternative splicing isoform of tau protein kinase I/glycogen synthase kinase 3 beta. J Neurochem 81:1073–1083

3. Plyte SE, Hughes K, Nikolakaki E, Pulve BJ, Woodgett JR (1999) Glycogen synthase kinase 3 is a potential drug target for african trypanosomiasis therapy. Biochem Biophys Acta 1114:147–162

4. Ali A, Hoeflich KA, Woodgett JR (2001) Glycogen synthase kinase-3: Properties, functions, and regulation. Chem Rev 101:2527–2540

5. Moon RT, Kohn RD, De Ferrari GV, Kaykas A (2004) WNT and beta-catenin signalling: diseases and therapies. Nat Rev Genet 5:691–701

6. Patel S, Doble B, Woodgett JR (2004) Glycogen synthase kinase-3 in insulin and Wnt signalling: a double-edged sword? Biochem Soc Trans 32:803–808

7. Kim KH, Song MJ, Yoo EJ, Choe SS, Park SD, Kim KB (2004) Regulatory role of glycogen synthase kinase 3 for transcriptional activity of ADD1/SREBP1c. J Biol Chem 279:5199–52006

8. Ross SE, Erickson RL, Hemati N, Macdougald OA (1999) Glycogen synthase kinase 3 is an insulin regulated C/EBPα kinase. Mol Cell Biol 19:8433–8441

9. Panchez JF, Sniderhan LF, Williamson AL, Fan S, Chakraborty-Sett S, Maggirwar SB (2003) Glycogen synthase kinase 3β-mediated apoptosis of primary cortical astrocytes involves inhibition of nuclear factor κB signaling. Mol Cell Biol 23:4649–4662

10. Lucas JJ, Hernandez F, Gomez-Ramos P, Moran MA, Hen R, Avila J (2001) Decreased nuclear δ-catenin, tau hyperphosphorylation and neurodegeneration in GSK-3β conditional transgenic mice. EMBO J 20:27–39

11. Aghdam SY, Barger SW (2007) Glycogen synthase kinase-3 in neurodegeneration and neuroprotection: lessons from lithium. Curr Alzheimer Res 4:21–31

12. Eldar-Finkolman H (2002) Glycogen synthase kinase 3: an emerging therapeutic target. Trends Mol Med 8:126–132

13. Hernandez F, Avila J (2008) The role of glycogen synthase kinase 3 in the early stages of Alzheimer's' disease. FEBS Lett 582:3848–3854

14. Dorronsoro I, Castro A, Martinez A (2002) Inhibitors of glycogen synthase kinase-3: future therapy for unmate medical needs? Expert Opin Ther Pat 12:1527–1536

15. Meijer L, Flajolet M, Greengard P (2004) Pharmacological inhibitors of glycogen synthase kinase 3. Trends Pharmacol Sci 25:471–480

16. Liang MH, Chuang DM (2006) Differential roles of glycogen synthase kinase-3 isoforms in the regulation of transcriptional activation. J Biol Chem 281:30479–30484

17. Davies SP, Reddy H, Caivano M, Cohen P (2000) Specificity and mechanism of action of some commonly used protein kinase inhibitors. Biochem J 351:95–105

18. Bertrand JA, Thieffine S, Vulpetti A, Cristiani C, Valsasina B, Knapp S, Kalisz HM, Flocco M (2003) Structural characterization of the GSK-3beta active site using selective and non-selective ATP-mimetic inhibitors. J Mol Biol 333:393–407

19. Lio JLJ (2007) Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. J Med Chem 50:409–424

20. Ghose AK, Herbertz T, Pippin DA, Salvino JM, Mallamo JP (2008) Knowledge based prediction of ligand binding modes and rational inhibitor design for kinase drug discovery. J Med Chem 51:5149–5171

21. Engh RA, Girod A, Kinzel V, Huber R, Bossemeyer D (1996) Crystal structures of catalytic subunit of cAMP-dependent protein kinase in complex with isoquinolinesulfonyl protein kinase inhibitors H7, H8, and H89 structural implications for selectivity. J Biol Chem 271:26157–26164

22. Lochner A, Moolman JA (2006) The many faces of H89: a review. Cardiovasc Drug Rev 24:261–274

23. Wang Z, Canagarajah BJ, Boehm JC, Kassisa S, Cobb MH, Young PR, Abdel-Meguid S, Adams JL, Goldsmith EJ (1998) Structural basis of inhibitor selectivity in MAP kinases. Structure 6:1117–1128

24. Lee JC, Laydon JT, McDonnell PC, Gallagher TF, Kumar S, Green D, McNulty MJ, Blumenthal JR, Keys SW, JE Land vatter, Strickler MM, McLaughlin IR, Siemens SM, Fisher GP, Livi JR, White JL, Younget A, Younget PR (1994) A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. Nature 372:739–746

25. Cuenda A, Rouse J, Doza YN, Meier R, Cohen P, Gallaghaer TF, Young PR, Lee JC (1994) SB 203580 is a specific inhibitor of a MAP kinase homologue which is stimulated by cellular stresses and interleukin-1. FEBS Lett 364:229–233

26. Thurmond RL, Wadsworth SA, Schafer PH, Zivin RA, Siekierka JJ (2001) Kinetics of small molecule inhibitor binding to p38 kinase. Eur J Biochem 268:5747–5754

27. Jacobs M, Hayakawa K, Swenson L, Bellon S, Fleming M, Taslimi P, Dora J (2006) The structure of dimeric ROCK I reveals the mechanism for ligand selectivity. J Biol Chem 281:260–268

28. Ishizaki T, Uehata M, Tamechika I, Keel J, Nonomura K, Maekawa M, Narumiya S (2000) Pharmacological properties of Y-27632, a specific inhibitor of Rho-associated kinases. Mol Pharmacol 57:976–983

29. Breitenlechner C, Gaßel M, Hidaka H, Kinzel V, Huber R, Engh RA, Bossemeyer D (2003) Protein kinase A in complex with Rho-kinase inhibitors Y-27632, Fasudil, and H-1152P: structural basis of selectivity. Structure 11:1595–1607

30. Komander D, Kular GS, Schuttelkopf AW, Deak M, Prakash KRC, Bain J, Elliott M, Garrido-Franco M, Kozikowski AP, Alessi DR, van Aalten DMF (2004) Interactions of LY333531 and other bisindolyl maleimide inhibitors with PDK1. Structure 12:215–226

31. Page SC, Bates PA (2006) Can MM-PBSA calculations predict the specificities of protein kinase inhibitors? J Comput Chem 27:1990–2007

32. 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

33. Reddy MR, Erion MD (2001) Free energy calculations in rational drug design, Springer-Verlag; ISBN: 978-0-306-46676-2

34. Frenkel D, Smit B, Understanding 2002 molecular simulation: from algorithm to applications. Academic, San Diego, CA. ISBN 0-12-267351-4

35. Leach AR (2001) Molecular modelling: principles and applications. Prentice-Hall, Upper Saddle Rvr, NJ. ISBN-978-0582382107

36. Honndorf VS, Coudevylle N, Laufer S, Becker S, Griesinger C (2008) Dynamics in the p38α MAP Kinase-SB203580 complex observed by liquid state NMR. Angew Chem Int Ed 47:3548–3551

37. Breitenlechner C, Gaßel M, Hidaka H, Kinzel V, Huber R, Engh RA, Bossemeyer D (2003) Protein kinase A in complex with Rho-kinase inhibitors Y-27632, Fasudil, and H-1152P: structural basis of selectivity. Structure 11:1595–1607

38. AbdulHameed MDM, Hamza A, Zhan CG (2006) Microscopic modes and free energies of 3-phosphoinositide-dependent kinase-1 (PDK1) binding with celecoxib and other inhibitors. J Phys Chem B 110:26365–26374

39. Carlsson J, Boukharta L, Aqvist J (2008) Combining docking, molecular dynamics, and the linear interaction energy method to predict binding modes and affinities for non-nucleoside inhibitors to HIV-1 reverse transcriptase. J Med Chem 51:2648–2656

40. Wang J, Morin P, Wang W, Kollman PA (2001) Combining docking, molecular dynamics, and the linear interaction energy method to predict binding modes and affinities for non-nucleoside inhibitors to HIV-1 reverse transcriptase. J Am Chem Soc 123:5221–5230

41. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein-ligand docking using GOLD. Proteins 52:609–623

42. Srivastava HK, Chourasia M, Kumar D, Sastry GN (2011) Comparison of computational methods to model DNA minor groove binders. J Chem Inf Model. doi:10.1021/ci100474

43. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2006) AMBER 9. University of California, San Francisco

44. Thomson DC, Humblet C, Joseph-McCarthy D (2008) Investigation of MM-PBSA rescoring of docking poses. J Chem Inf Model 45:1081–1091

45. 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

46. Jones G, Willett P, Glen RC (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J Mol Biol 245:43–53

47. Verdonk ML, Chessari G, Cole JC, Hartshorn MJ, Murrary CW, Nissink JWM, Taylor RD, Taylor R (2005) Modeling water molecules in protein − ligand docking using GOLD. J Med Chem 48:6504–6515

48. Jorgensen WL, Chandrasekhar J, Madura JD, Klein MLJ (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935

49. Essmann U, Perera L, Berkowitz ML, Darden TA, Lee H, Pederson LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577S–8593S

50. Ryckaert JP, Cicotti G, Berendsen HC (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

51. Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem 23:1623–1641

52. Wang J, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model 25:247–260

53. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25:1157–1174

54. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman PA (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24:1999–2012

55. Wong S, Amaro RE, McCammon JA (2009) MM-PBSA captures key roles of intercalating water molecules at a protein-protein interface. J Chem Theor Comput 5:422–429

56. Strockbine B, Rizzo RC (2007) Binding of anti-fusion peptides with HIVgp41 from molecular dynamics simulations: quantitative correlation with experiment. Proteins 67:630–642

57. Connolly ML (1983) Analytical molecular surface calculation. J Appl Cryst 16:548–558

58. Gilson MK, Zhou HX (2007) Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct 36:21–42

59. Mobley DC, Dill KA (2009) Binding of small-molecule ligands to proteins: "what you see" is not always "what you get. Structure 17:489–498

60. Perakyla M, Nordman N (2001) Energetic analysis of binding of progesterone and 5β-androstane-3,17-dione to anti progesterone antibody DB3 using molecular dynamics and free energy calculations. Prot Eng 14:753–758

61. 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 TE (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Accts Chem Res 33:889–897

62. Wang J, Morin P, Wang W, Kollman PA (2001) Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J Am Chem Soc 123:5221–5230

63. Ferrari AM, Degliesposti G, Sgobba M, Rastelli G (2007) Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. Bioorg Med Chem 15:7865–7877

64. Laitinen T, Rouvinen J, Peräkylä M (2003) MM-PBSA free energy analysis of endo-1,4-xylanase II (XynII)–substrate complexes: binding of the reactive sugar in a skew boat and chair conformation. Org Biomol Chem 1:3535–3540

65. Hall-Jackson CA, Goedert M, Hedge P, Cohen P (1999) Effect of SB 203580 on the activity of c-Raf in vitro and in vivo. Oncogene 18:2047–2054

66. Badger J, Minor I, Kremer MJ, Oliveira MA, Smith TJ, Griffith JP, Guerin DM, Krishnaswamy S, Luo M, Rossmann MG, Mckinlay MA, Diana GD, Dutko FJ, Fancher M, Rueckert RR, Heinz BA (1988) Structural analysis of a series of antiviral agents complexed with human rhinovirus 14. Proc Natl Acad Sci USA 85:3304–3308

67. Lewis PJ, Jonge M, Daeyaert F, Koymans L, Vinkers M, Heeres J, Janssen PAJ, Arnold E, Das K, Clark AD, Hughes SH, Boyer PL, de B'thune MP, Pauwels R, Andries K, Kukla M, Ludovici D, Corte BD, Kavash R, Ho C (2003) On the detection of multiple-binding modes of ligands to proteins, from biological, structural, and modeling data. J Comput Aided Mol Des 17:129–134

68. Engler TA, Henry JR, Malhotra S, Cunningham B, Furness K, Brozinick J, Burkholder TP, Clayton J, Diefenbacher C, Hawkins E, Iversen PW, Li Y, Lindstrom TD, Marquart AL, McLean J, Mendel D, Misener E, Briere D, O'Toole JC, Porter WJ, Queener S, Reel JK, Owens RA, Brier RA, Eessalu TA, Wagner JR, Campbell RA, Vaughn RS (2004) Substituted 3-imidazo[1,2-a]pyridin-3-yl- 4-(1,2,3,4-tetrahydro-[1, 4]diazepino-[6,7,1-hi]indol-7-yl)pyrrole-2,5-diones as highly selective and potent inhibitors of glycogen synthase kinase-3. J Med Chem 47:3934–3937

69. Zhang HC, Bonaga LVR, Ye H, Derian CK, Damiano BP, Maryanoff BE (2007) Novel bis(indolyl)maleimide pyridino-phanes that are potent, selective inhibitors of glycogen synthase kinase-3. Bioorg Med Chem Letts 17:2863–2868

70. Komander D, Kular GS, Bain J, Elliott M, Alessi DR, van Aalten DMF (2003) Structural basis for UCN-01 (7-hydroxystaurosporine) specificity and PDK1 (3-phosphoinositide-dependent protein kinase-1) inhibition. Biochem J 375:255–262

71. Bhat R, Xue Y, Berg S, Hellberg S, Ormö M, Nilsson Y, Radesäter AC, Jerning E, Markgren PO, Borgegård T, Nylöf M, Giménez-Cassina A, Hernández F, Díaz-Nido J, Avila J (2003) Structural insights and biological effects of glycogen synthase kinase 3-specific inhibitor AR-A014418. J Biol Chem 278:45937–45945

72. Khanfar MA, Hill RA, AKaddoumi A, El Sayed KA (2010) Discovery of novel GSK-3β inhibitors with potent in vitro and in vivo activities and excellent brain permeability using combined ligand- and structure-based virtual screening. J Med Chem 53:8534–8545

73. Zhang N, Zhong R, Yan H, Jiang YC (2010) Structural features underlying selective inhibition of GSK3β by Dibromocantharelline: implications for rational drug design. Biol Drug Design. doi:10.1111/j.1747-0285.2010.01069.x

74. Babu PA, Chitti S, Rajesh B, Prasanth VV, RadhaKishen JV, Vali RK (2010) In silico based ligand design and docking studies of GSK-3ß inhibitors. Chem Bioinf J 10:1–12

Supporting information available

Supporting information contains superposition of all docked conformations of H89, SB203580, Y27632, BIS-3, BIS-8 and UCN1 ligands, superposition of GSK-3β-BIS-3 and PDK1-BIS-3 complexes, structure of the GSK-3β-BIS-8 (rank-9) complex, superposition of GSK-3β-STU and GSK-3β-UCN1 complexes and superposition of GSK-3β-UCN1 and PDK1-UCN1 complexes.

ORIGINAL PAPER

Intermolecular interactions between gold clusters and selected amino acids cysteine and glycine: a DFT study

Hu-Jun Xie · Qun-Fang Lei · Wen-Jun Fang

Abstract The intermolecular interactions between Au_n (n= 3–4) clusters and selected amino acids cysteine and glycine have been investigated by means of density functional theory (DFT). Present calculations show that the complexes possessing $Au-NH_2$ anchoring bond are found to be energetically favored. The results of NBO and frontier molecular orbitals analysis indicate that for the complex with anchoring bonds, lone pair electrons of sulfur, oxygen, and nitrogen atoms are transferred to the antibonding orbitals of gold, while for the complex with the nonconventional hydrogen bonds ($Au\cdots H-O$), the lone pair electrons of gold are transferred to the antibonding orbitals of O-H bonds during the interaction. Furthermore, the interaction energy calculations show that the complexes with $Au-NH_2$ anchoring bond have relatively high intermolecular interaction energy, which is consistent with previous computational studies.

Keywords Amino acids · DFT calculations · Gold clusters · Interaction energies

Introduction

Among the various nanomaterials, gold nanomaterials have attracted much attention due to possessing good biocom-

patibility, facile synthesis and conjugation to lots of biomolecular moieties, and hence are widely applied in chemistry, materials, biological and medical science [1–5]. The high affinities of gold nanoparticles for biomolecules have extensive applications in modern protein engineering and gene technology [6–8]. Also, the formation of highly ordered molecular self-assembled networks of amino acids on the well-controlled metal surfaces have been previously studied [9–12]. However, very little is known about the mechanisms underlining the formation of these layers. A description of these mechanisms at the atomic level is currently lacking and deserves special attention. Therefore, the study on intermolecular interactions between Au_n (n= 3–4) clusters and selected amino acids cysteine and glycine, is an excellent framework to understand the biomolecule adsorption on the Au surface.

Relevant studies have been carried out to explore the interactions between amino acids and gold surface experimentally and theoretically [13–15]. Ulstrup et al. have investigated the cysteine deposited on the Au(111) surface using the scanning tunneling microscopy (STM). The results reveal that novel network-like cluster structures of the layers for both cysteine and cystine on the Au surface are found, each cluster is shown to include six cysteine or three cystine molecules. This super-structural feature is most likely due to the formation of hydrogen bonds among adsorbed molecules, which is further supported by the observations of alkanethiol under the same experimental conditions. Intermolecular and intramolecular interactions of the charged hydrophilic ammonium and carboxylic groups in cysteine or cystine are thus crucial for the cluster formation [16, 17]. Canepa and Lavagnino [18] have reported on a metastable deexcitation spectroscopy investigation of the growth of L-cysteine layers deposited under UHV conditions on well-defined Au(110)-(1×2) and Au(111) surfaces. The different growth mode was tentatively assigned to the added rows of the reconstructed Au(110) surface, which behaves as

H.-J. Xie (✉)
Department of Applied Chemistry,
Zhejiang Gongshang Universty,
Hangzhou 310035, China
e-mail: hujunxie@gmail.com

Q.-F. Lei · W.-J. Fang
Department of Chemistry, Zhejiang University,
Hangzhou 310027, China

Q.-F. Lei
e-mail: qflei@zju.edu.cn

extended defects effectively promoting the formation of the Au-S bond. De Renzi et al. [19] have studied the ultrahigh vacuum adsorption of cysteine layers on the Au(111) surface by X-ray photoelectron (XPS) and high-resolution energy loss spectroscopies (HREELS). The experiments show that at room temperature the molecule is mainly weakly adsorbed on the surface via at low coverage chemisorption on the step edge, while at 330 K chemisorption through S-Au bonding prevails. The results from spectroscopic evidence substantially confirms and clarifies the mechanism of cysteine adlayer formation on the Au(111) surface. In addition, the red-shift of H vibrational modes indicate a coverage dependence of the H-mode frequency, which clearly supports its intermolecular origin. This finding is a good example of the extreme sensitivity of low-frequency vibrational modes to the details of molecule-molecule interactions.

On the other hand, relevant theoretical works have been devoted to the studies of geometries, bonding strengths, energies and molecule orbital properties during the interaction between the Au surfaces and different amino acids. Density functional theory (DFT) calculations of cysteine adsorption on Au(100) by Höffling and Ortmann [20] have shown that flat adsorption geometries are energetically favored over upright configurations, taking advantage of both amino and thiolate bonds. Mateo-Marti et al. [21] have investigated the first stages leading to the formation of self-assembled monolayers of S-cysteine molecules adsorbed on the Au(111) surface. The DFT calculations for the adsorption of individual cysteine molecules on Au(111) show low-energy barriers all over the 2D Au(111) unit cell.

Understanding the adsorption, bonding, and interaction of the simplest constituents of proteins (i.e., amino acids) on the Au surfaces is a necessary step toward broad applications in the interdisciplinary emerging field of nanobiotechnology and biotechnology. This study aims at elucidating the main mechanisms for the amino acids adsorption on the Au surfaces. For this reason, in the present work, the intermolecular interactions between Au_n (n=3–4) clusters and selected amino acids cysteine and glycine have been investigated theoretically by DFT calculations. Special emphasis is put on the analysis of geometry configurations, electronic structures and accompanying electron transfers of selected amino acids and Au cluster complexes during the interaction.

Computational details

In the present calculations, the Au_n (n=3–4) clusters were served as simple model for the Au surface, and geometries of Au_n·Cysteine and Au_n·Glycine complexes were optimized with the density functional theory (DFT) method using the Becke's three-parameter Lee–Yang–Parr exchange-correlation functional (B3LYP) [22, 23]. The standard 6-311++g(d,p)

basis set was used for all atoms in cysteine and glycine. The relativistic effective core potential RECP developed by Ross et al. [24] with the primitive basis set (5s5p4d) combined with Los Alamos LANL2DZ RECP by Hay and Wadt [25, 26] were used for the Au_n cluster. The reliability of this method has been verified in previous studies involved the calculations of Au-ligand systems [27–32]. Then frequency calculations were employed to confirm the structures as minimum points in energy. The interaction energies, complexation enthalpies and Gibbs free energies at 298 K were calculated for each optimized structure by vibrational analysis with ZPE correction. To obtain an insight into the charge distribution of the intermolecular interactions, the geometries obtained from DFT calculations were used to perform the NBO analysis [33–36]. All calculations were performed using the Gaussian 09 program [37].

Results and discussion

Geometry and relative energy of the complexes

On the basis of extensively computational studies on the complexes of the Au_n·Cysteine and Au_n·Glycine (n=3,4), 12 possible geometry conformations have been proposed as shown in Fig. 1. The initial geometries of Au_n·Cysteine and Au_n·Glycine complexes for optimization were generated by placing gold clusters near the active sites of glycine and cysteine. The active sites of two amino acids are amine, carboxylic, and sulfur groups. These groups have electron-rich nitrogen, oxygen and sulfur atom, forming the anchoring bonds with gold clusters in the complexes, donating electron density to the 5 d and 6 s orbitals of Au via their lone pairs. Gold can also play the role as a proton acceptor and form nonconventional hydrogen bonds with hydroxide (Au···H–O). These two factors are responsible for the stability of all complexes. The interaction between amino acids and gold clusters is either monodentate or bidentate, where in the latter case it usually involves non-conventional hydrogen bonding.

As for Au_n·Cysteine complexes, cysteine is able to anchor the gold clusters through its amine, carboxylic, and sulfur groups, and yields the Au-N, Au-O and Au-S anchoring bonds. As shown in Table 1, the relative energy of Au_3·Cysteine complexes spread over a range of 8.7 kcal mol^{-1}. The complex-1 and complex-4 have the lowest energy in Au_3·Cysteine and Au_4·Cysteine isomers, respectively. The relatively high stability can be ascribed to strong interaction between the Au_n clusters and nitrogen atom in cysteine. By contrast the complex-3 and complex-6 have relatively high energy in Au_3·Cysteine and Au_4·Cysteine isomers, respectively. In addition, the optimized structures of the Au_n·Cysteine complexes are displayed in Fig. 1. For complex-1, the average Au-Au distance is equal to 2.768 Å,

and the Au_1-N bond length is found to be 2.220 Å, which is slightly shorter than previous calculated value of 2.260 Å from PW91/Plane-Wave-Basis method [20]. The Au_1-N bond length approaches the sum of the covalent radii of Au (1.35 Å) and N (0.750 Å) [38], indicating a strong bond. Vibrational frequency analysis is a powerful technique that can be used to reveal information on adsorption configuration and bond strengths, which has been validated by previous studies [39–42]. As demonstrated in Table 2, in contrast to isolated cysteine, the N-H bond in complex-1 has been increased by 0.004 Å after interaction with Au_3 cluster. Furthermore, it is clear that the stretch mode of ν(N-H) undergoes a red shift by 48 cm^{-1} with respect to

Table 1 The relative energy (E in kcal mol^{-1}) of the Au_n·Cysteine and Au_n·Glycine complexes

	Complex	E		Complex	E
Au_3·Cysteine	1	0.0	Au_3·Glycine	7	0.0
	2	0.2		8	11.7
	3	8.7		9	7.6
Au_4·Cysteine	4	0.0	Au_4·Glycine	10	0.0
	5	1.1		11	12.2
	6	5.4		12	8.1

Table 2 The vibration frequency (υ in cm^{-1}) and the variation of the bond lengths (Δr in Å) and the corresponding frequency shifts ($\Delta\upsilon$ cm^{-1}) of the selected bonds obtained at the B3LYP/6-311++g(d,p) and Lanl2dz level

Bond	Complex	Δr	υ	$\Delta\upsilon$	Complex	Δr	υ	$\Delta\upsilon$
C = O	1	0.001	1820	1	2	0.018	1740	−79
N-H		0.004	3560	−48		0.001	3618	10
O-H		0.001	3743	2		0.001	3727	−14
S-H		0.001	2691	−2		0.012	2522	−171
C = O	3	−0.001	1823	4	4	0.001	1821	2
N-H		−0.001	3627	19		0.004	3562	−46
O-H		0.010	3541	−200		0.001	3743	2
S-H		0.003	2670	−23		0.001	2691	−2
C = O	5	−0.002	1827	8	6	0.026	1702	−117
N-H		−0.001	3626	18		−0.001	3623	15
O-H		0.005	3648	−93		0.021	3304	−437
S-H		0.003	2672	−21		0.001	2688	−5
C = O	7	0.001	1821	2	8	0.015	1747	−72
N-H		0.003	3565	−33		0.001	3506	−92
O-H		0.001	3748	−2		0.001	3735	−15
C = O	9	0.025	1703	−116	10	0.001	1821	2
N-H		0.000	3612	14		0.003	3566	−32
O-H		0.020	3320	−430		0.001	3748	−2
C = O	11	0.014	1753	−66	12	0.024	1704	−115
N-H		−0.001	3620	22		0.001	3611	13
O-H		0.001	3737	−13		0.001	3297	−453

that of the uncoordinated group. The next stable species is complex-2, it is slightly less stable than complex-1 only by 0.02 kcal mol^{-1}. The predicted average Au-Au distance is 2.771 Å, slightly longer than that of complex-1. The Au$_1$-S distance is found to be 2.449 Å, very close to the sum of the covalent radii 2.360 Å of sulfur (1.020 Å) and gold (1.340 Å) [38]. The result is in good agreement with a previous calculated value of 2.450 Å [20, 43]. Thus, one may expect a strong covalent contribution to the thiolate bonding. As shown in Table 2, the complex-2 displays relatively obvious red shifts of selected bonds in contrast to the isolated cysteine. Such as the frequency shifts of S-H bond is reduced by 171 cm^{-1}. The highest-energy structure among the Au$_3$·Cysteine complexes in Table 1 is complex-3. As Fig. 1 displays, the average Au-Au distance is equal to 2.780 Å, significantly longer than that of complex-1 and complex-2. The distance of anchoring bond Au$_1$-O in the complex-3 is predicted to be 2.256 Å. It is noted that the stretch frequency of υ(O-H) is markedly reduced by 200 cm^{-1} compared to the isolated cysteine. For Au$_4$·Cysteine complexes, the relative stability of complexes is complex-4> complex-5>complex-6, they are corresponding to the binding modes between Au$_4$ cluster with N, S and O atoms in cysteine, the stability order is similar to that of Au$_3$·Cysteine complexes. As Table 2 shows, the stretch frequency of υ(S-H) in complex-6 shows an obvious red shift by 437 cm^{-1}.

For Au$_n$·Glycine complexes, the Au$_n$ cluster can bind glycine via monodentate or bidentate interaction with N, O

(C = O), or H(O-H) atoms. As Table 1 shows, complex-7 and complex-10 have relatively low energy in Au$_3$·Glycine and Au$_4$·Cysteine isomers, respectively. The relatively high stability is ascribed to the strong interaction between the Au$_n$ clusters and amino group in glycine. This is agree with the interaction between the Au$_n$ clusters and cysteine. As displayed in Fig. 1, the average Au-Au distance in complex-7 is found to be 2.770 Å, and the Au$_1$–N bond length is equal to 2.212 Å. As Table 2 displays, the distance of N-H bond is lengthened by 0.003 Å, the corresponding stretch frequency υ(N-H) goes through a red shift by 33 cm^{-1}. Complex-9 has bidentate interaction via Au–O and Au···H–O binding modes [44, 45], which is less stable than complex-1 by 7.6 kcal mol^{-1}. The average Au-Au distance is predicted to be 2.781 Å, slightly longer than that of complex-7. The stretch frequency υ(O-H) occurs obvious red shift by 430 cm^{-1} in contrast to single glycine. For Au$_4$·Glycine complexes, the stability order of complexes is complex-10>complex-12>complex-11.

Natural population analysis and vertical transition energies

The natural population analysis (NPA) charges with the corresponding variations in the related atoms and the total transferred charges of Au$_n$·Cysteine and Au$_n$·Glycine complexes are listed in Table 3. Figure 2 presents the frontier molecular orbitals of Au$_n$·Cysteine and Au$_n$·Glycine complexes associated with charge-transfer analysis. As

Table 3 The NPA charges (a.u.) with the corresponding variations in the related atoms and the total transferred charges for 12 $Au_n \cdot$Cysteine and $Au_n \cdot$Glycine complexes

		e_{mono}[a]	e_{com}[b]	Δe[c]	e_t[d]			e_{mono}	e_{com}	Δe	e_t
1	Au_1	−0.07	0.23	0.31	−0.14	8	Au_1	−0.07	0.25	0.33	−0.06
	N	−0.89	−0.92	−0.03			O	−0.62	−0.68	−0.07	
2	Au_1	−0.07	0.19	0.26	−0.17	9	Au_1	−0.07	0.28	0.35	−0.03
	S	−0.06	0.03	0.09			Au_3	0.04	−0.21	−0.25	
3	Au_1	−0.07	0.26	0.33	−0.03		O	−0.62	−0.69	−0.07	
	O	−0.62	−0.70	−0.08			H	0.52	0.50	−0.01	
4	A_1	−0.06	0.32	0.38	−0.14	10	Au_1	−0.06	0.32	0.38	−0.15
	N	−0.89	−0.92	−0.03			N	−0.89	−0.92	−0.02	
5	Au_1	−0.06	0.26	0.32	−0.19	11	Au_1	−0.06	0.34	0.40	−0.07
	S	−0.06	0.04	0.10			O	−0.62	−0.67	−0.06	
6	Au_1	−0.06	0.18	0.24	−0.03	12	Au_1	−0.06	0.36	0.42	−0.02
	O	−0.62	−0.69	−0.07			Au_3	0.11	−0.34	−0.45	
7	Au_1	−0.07	0.23	0.31	−0.15		O	−0.62	−0.68	−0.07	
	N	−0.89	−0.92	−0.02			H	0.52	0.40	−0.12	

[a] e_{mono} refers the atomic charge of monomer

[b] e_{com} means the atomic charge of complex

[c] Δe is evaluated as the NPA charge difference between the atomic charge in corresponding complexes and monomers

[d] e_t is defined as the total NPA charge transfer from amino acids (cysteine and glycine) to Au_n (n=3–4)

demonstrated in Table 3, evidently in the case of all complexes described in this paper, the extra electron is localized on the gold clusters. The charge distribution of the complexes closely coincides with the electron affinity (EA) values. The EA of the gold clusters (3.85, and 2.77 eV for Au_3, and Au_4, respectively) is greater than the EA of the cysteine (−0.39 eV) and glycine (−0.51 eV) molecules [46], and for this reason, the extra electron is localized on the

Fig. 2 The frontier molecular orbitals of the Aun·Cysteine and Aun·Glycine complexes associated with charge-transfer analysis

1-HOMO 1-LUMO 2-HOMO 2-LUMO 3-HOMO 3-LUMO

4-HOMO 4-LUMO 5-HOMO 5-LUMO 6-HOMO 6-LUMO

7-HOMO 7-LUMO 8-HOMO 8-LUMO 9-HOMO 9-LUMO

10-HOMO 10-LUMO 11-HOMO 11-LUMO 12-HOMO 12-LUMO
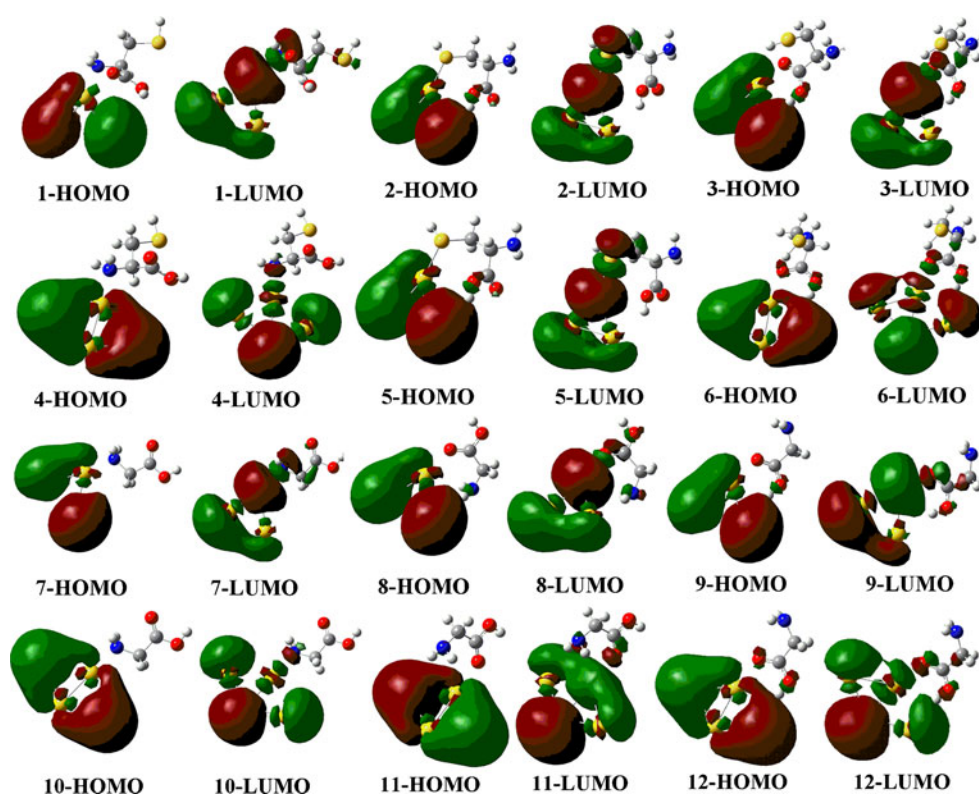
gold clusters. Thus, in complex-1, the charge is transferred from cysteine to Au_3 with the value of 0.14. In the monomer involved the Au_3 cluster and cysteine, the charge populations on the Au_1 and N atoms are predicted to be -0.07 and -0.89, respectively. While in the complex, it is clear that the charge of Au_1 atom has been increased by 0.31, and the charge of N atom has been reduced by 0.03. The charge populations of other complexes are similar to that of complex-1. As Table 3 displays, the calculated results indicate that the gold atom (Au_1) in the complexes that is bonded near to the electron rich sites of the amino acids is positively charged. This is in well agreement with previous theoretical values determined by Kumar and co-workers [47].

On the basis of frontier molecular orbitals analysis as presented in Fig. 2, it is clear that for the complex with Au–S, Au–NH_2 and Au–O anchoring bonds, lone pair electrons of sulfur, oxygen, and nitrogen are transferred to the antibonding orbitals of the gold. For example, as listed in Table 3, complex-5 has the charge transfer from cysteine to Au_3 with the value of 0.19, and complex-10 shows similar charge populations with complex-5. For complex-9 and complex-12 possessing nonconventional hydrogen bonds (Au···O-H), the charge is transferred from the lone pair of Au to the σ^*_{O-H} orbitals. Table 4 exhibits the predicted vertical transition energies of the Au_n·Cysteine and Au_n·Glycine complexes. Complex-8 has the lowest vertical transition energies of 1.151 eV among all complexes, it corresponds to the maximum absorption wavelength of 1077.43 nm, excited from the HOMO to LUMO as shown in Fig. 2. While complex-10 has the highest vertical transition energies of

Table 4 The predicted vertical transition energies (ΔE in eV), maximum absorption wavelength (λ in nm), oscillator strengths (f) of the Au_n·Cysteine and Au_n·Glycine complexes

Complex	ΔE	λ	f[a]	Transition[b]
1	1.334	929.51	0.0119	61→62
2	1.453	853.58	0.0143	61→62
3	1.354	915.60	0.0081	61→62
4	1.632	760.01	0.0071	70→71
5	1.634	758.92	0.0074	70→71
6	1.553	798.55	0.0072	70→71
7	1.379	899.18	0.0116	49→50
8	1.151	1077.43	0.0103	49→50
9	1.365	908.29	0.0080	49→50
10	1.638	757.10	0.0087	58→59
11	1.504	824.16	0.0030	58→59
12	1.558	796.05	0.0075	58→59

[a] The magnitude of f is in proportion to that of absorption strength
[b] Transition means the electron promotion from HOMO to LUMO (see Fig. 2)

Table 5 The predicted intermolecular interaction energy (ΔE in kcal mol^{-1}) between Au_n (n=3–4) and selected amino acids cysteine and glycine

Complex	ΔE	ΔH	ΔG	Complex	ΔE	ΔH	ΔG
1	16.2	15.9	5.1	7	19.9	19.6	9.3
2	16.0	15.8	3.5	8	8.1	7.6	−2.8
3	7.5	7.0	−4.6	9	12.3	11.9	0.4
4	16.7	16.4	5.0	10	20.2	20.0	9.0
5	15.6	15.3	2.8	11	7.4	7.1	−3.0
6	11.4	11.0	−1.0	12	12.1	11.7	0.2

1.638 eV, it corresponds to the maximum absorption wavelength of 757.10 nm, excited from the HOMO to LUMO.

Interaction energies

The interaction energies (ΔE) are defined as the energy difference between the energy summation of two individually optimized monomers and the fully optimized complexes.

$$\Delta E = E_{Au-clusters} + E_{amino-acid} - E_{complex}$$

Table 5 presents the predicted intermolecular interaction energy of all complexes. In the case of Au_3·Cysteine complexes, the predicted interaction energy amounts to 16.2 kcal mol^{-1} in complex-1, which agrees well with a previous calculated value of 16.5 kcal mol^{-1} obtained from PW91/Plane-Wave-Basis method [20]. Complex-1 is the energetically most-favored structure, the relatively high stability can be ascribed to strong interaction between the N atom and Au_3 cluster. Furthermore, the next stable specie is complex-2, the interaction energy is less stable than complex-1 only by 0.2 kcal mol^{-1}. In the case of Au_4·Cysteine complexes, complex-4 has the largest interaction energies of 16.7 kcal mol^{-1}. The association of Au_3 cluster with thiol in cysteine is relatively weak with the interaction energy of 16.0 kcal mol^{-1}, while complex-3 related to the interaction between Au_3 cluster and O atom in the cysteine has the smallest interaction energy of 7.5 kcal mol^{-1}. The calculations indicate that the Au_3·Cysteine complexes with Au-NH_2 anchoring bond are found to be energetically favored. This is in good agreement with available theoretical values at the PW91/ Plane-Wave-Basis level of theory [20].

In the case of the Au_3·Glycine complexes, complex-7 possesses relatively large interaction energies of 19.9 kcal mol^{-1}. While for the Au_4·Glycine complexes, complex-10 has relatively large interaction energy of 20.2 kcal mol^{-1}. The calculations show that the Au_3·Glycine complexes with

Au-NH$_2$ anchoring bond are relatively stable, which is consistent with previous computational studies [20].

Conclusions

In this work, density functional theory has been used to explore geometry configurations, electronic structures, charge populations and interaction energies of Au$_n$·Cysteine and Au$_n$·Glycine complexes. The calculations show that complex-1 and complex-4 are the most energetically favorable in the Au$_3$-cysteine and Au$_4$-cysteine complexes. For the Au$_n$-Glycine complexes, complex-7 and complex-10 are the most energetically favorable in the Au$_3$-Glycine and Au$_4$-Glycine complexes. In addition, the results of the NBO analysis indicate that the complex with anchoring bonds, lone pair electrons of sulfur, oxygen, and nitrogen are transferred to the antibonding orbitals of gold, while for the nonconventional hydrogen bonds, the lone pair electrons of gold are transferred to the antibonding orbitals of O-H bonds during the interaction. Meanwhile, the interaction energies calculations indicate that the Au-NH$_2$ anchoring bond acts as the dominate factor in stabilizing these complexes. Present results provide a basis for understanding the amino acids or proteins absorption on the Au surface at the atomic level, since in those systems the NH$_2$ or S-H groups are usually engaged in peptide bonds along the primary chain.

References

1. Mitchell GP, Mirkin CA, Letsinger RL (1999) J Am Chem Soc 121:8122–8123
2. Cao YWC, Jin R, Mirkin CA (2002) Science 297:1536–1540
3. Cavalleri O, Gonella G, Terreni S, Vignolo M, Floreano L, Morgante A, Canepa M, Rolandi R (2004) Phys Chem Chem Phys 6:4042–4046
4. Duan S, Wu DY, Xu X, Luo Y, Tian ZQ (2010) J Phys Chem C 114:4051–4056
5. Lazarus LL, Yang ASJ, Chu S, Brutchey RL, Malmstadt N (2010) Lab Chip 10:3377–3379
6. Park SJ, Lazarides AA, Mirkin CA, Letsinger RL (2001) Angew Chem Int Edit 40:2909–2912
7. Katz E, Willner I (2004) Agnew Chem Int Ed 43:6042–6108
8. Lopez-Ramirez MR, Garcia-Ramos JV, Otero JC, Castro JL, Sanchez-Cortes S (2007) Chem Phys Lett 446:380–384
9. Nazmutdinov RR, Zhang JD, Zinkicheva TT, Manyurov IR, Ulstrup J (2006) Langmuir 22:7556–7567
10. Kühnle A, Linderoth TR, Besenbacher F (2006) J Am Chem Soc 128:1076–1077
11. Schillinger R, Sljivancanin Z, Hammer B, Greber T (2007) Phys Rev Lett 98:136102
12. Crespilho FN, Lima FCA, da Silva ABF, Oliveira ON Jr, Zucolottoc V (2009) Chem Phys Lett 469:186–190
13. Kühnle A, Linderoth TR, Besenbacher F (2003) J Am Chem Soc 125:14680–14681
14. Pakiari AH, Jamshidi Z (2007) J Phys Chem A 111:4391–4396
15. Sudeep PK, Joseph STS, Thomas KG (2005) J Am Chem Soc 127:6516–6517
16. Zhang JD, Chi QJ, Nielsen JU, Friis EP, Andersen JET, Ulstrup J (2000) Langmuir 16:7229–7237
17. Wang Y, Chi QJ, Hush NS, Reimers JR, Zhang JD, Ulstrup J (2009) J Phys Chem C 113:19601–19608
18. Canepa M, Lavagnino L, Pasquali L, Moroni R, Bisio F, De Renzi V, Terreni S, Mattera L (2009) J Phys Condens Matter 21:264005
19. De Renzi V, Rousseau R, Marchetto D, Biagi R, Scandolo S, Del Pennino U (2008) J Phys Chem C 112:14439–14445
20. Höffling B, Ortmann F, Hannewald K, Bechstedt F (2010) Phys Rev B 81:045407
21. Mateo-Marti E, Rogero C, Gonzalez C, Sobrado JM, de Andres PL, Martin-Gago JA (2010) Langmuir 26:4113–4118
22. Beek AD (1993) J Chem Phys 98:5648–5652
23. Lee C, Yang WT, Parr RG (1988) Phys Rev B 37:785–789
24. Ross RB, Powers JM, Atashroo T, Ermler WC, LaJohn LA, Christiansen PA (1990) J Chem Phys 93:6654–6670
25. Hay PJ, Wadt WR (1985) J Chem Phys 82:270–283
26. Hay PJ, Wadt WR (1985) J Chem Phys 82:299–310
27. Kryachko ES, Karpfen A, Remacle F (2005) J Phys Chem A109:7309–7318
28. Liu YX, Zhang DJ, Zhou JH, Liu CB (2010) J Phys Chem A 114:6164–6170
29. Kryachko ES, Remacle F (2005) J Phys Chem B 109:22746–22757
30. Kryachko ES, Remacle F (2007) J Chem Phys 127(194305):1–11
31. Lv G, Wei FD, Jiang H, Zhou YY, Wang XM (2009) J Mol Struct Theochem 915:98–104
32. Kryachko ES, Remacle F (2005) Nano Lett 5:935–739
33. Reed AE, Weinstock RB, Weinhold F (1985) J Chem Phys 83:735–746
34. Reed AE, Weinhold F (1985) J Chem Phys 83:1736–1740
35. Reed AE, Curtiss LA, Weinhold F (1988) Chem Rev 88:899–926
36. Reed AE, Schleyer PvR (1990) J Am Chem Soc 112:1434–1445
37. 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, Peralta JE Jr, 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.1. Gaussian Inc, Wallingford, CT
38. Sargent-Welch (1980) Table of Periodic Properties of the Elements. Sargent-Welch Scientific Company, Skokie, IL
39. Li YC, Yang CL, Sun MY, Li XX, An YP, Wang MS, Ma XG, Wang DH (2009) J Phys Chem A 113:1353–1359

40. Gronbeck H, Hellman A, Gavrin A (2007) J Phys Chem A 111:6062–6067
41. Grybos R, Benco L, Bucko T, Hafner J (2009) J Comput Chem 30:1910–1922
42. Ding XL, Li ZY, Yang JL, Hou JG, Zhu QS (2004) J Chem Phys 121:2558–2562
43. Di Felice R, Selloni A, Molinari E (2003) J Phys Chem B 107:1151–1156
44. Kryachko ES, Remacle F (2005) Chem Phys Lett 404:142–149
45. Shafai GS, Shetty S, Krishnamurty S, Shah V, Kanhere DG (2007) J Chem Phys 126:014704
46. Taylor KJ, Pettiette-Hall CL, Cheshnovsky O, Smalley RE (1992) J Chem Phys 96:3319–3327
47. Kumar A, Mishra PC, Suhai S (2006) J Phys Chem A 110:7719–7727

Protein-protein docking on molecular models of *Aspergillus niger* RNase and human actin: novel target for anticancer therapeutics

**Ravi Kumar Gundampati · Rajasekhar Chikati · Moni Kumari · Anurag Sharma ·
Daliparthy Devi Pratyush · Medicherla V. Jagannadham · Chitta Suresh Kumar ·
Mira Debnath Das**

Abstract The 3D models of human actin protein and *A. niger* RNase were designed using the templates ACTBIND (PDB ID: 3D3Z) and crystalline profilin-beta-actin (PDB ID: 2BTF), respectively in Modeller9v5. These models are testified using several validation methods including PRO-CHECK, ERRAT, WHAT-IF, PROSA2003 and VERIFY-3D. The stereo-chemical quality of the models was judged by Ramachandran plot with PROCHECK. The total quality G-factor −0.2, shows a good quality model. The ERRAT score for the human actin and A.niger RNase models are 86.104 and 84.615, respectively, fit well within the range of a high quality model. The ERRAT score for the templates 2BTF and 3D3Z are 91.111 and 97.391, respectively. The WHAT-IF evaluation justifies a reasonable homology model structure as none of the scores for each residue in the homology model is lower than −5.0. The energy-minimized model of human actin with PROSA reveals the Z-score value −10.52 between native conformations of the crystal structures. The VERIFY 3D average score is 0.36. All evidence suggests that the geometric quality of the backbone conformation, the residue interaction, the residue contact and the energy profile of the structures were well within the limits of reliable structures. The interaction energy of docking was calculated using the HEX server. The Etotal, lowest docked energy, and calculated RMSD values were −1.608 kcal mol^{-1}, -8.369 kcal mol^{-1} and 0.617Å, respectively. The study presented in the current project may be useful to design molecules that may have anticancer activity.

R. K. Gundampati · M. Debnath Das (✉)
School of Biochemical Engineering, Institute of Technology,
Banaras Hindu University,
Varanasi 221005, India
e-mail: debnathmira@yahoo.co.in

R. K. Gundampati
e-mail: ravi_33102000@yahoo.com

R. Chikati · C. S. Kumar (✉)
DBT-Bioinformatics Infrastructure Facility (BIF),
Department of Biochemistry, Sri Krishnadevaraya University,
Anantapur 515003, India
e-mail: chitta34c@gmail.com

M. Kumari · A. Sharma · M. V. Jagannadham
Molecular Biology Unit, Institute of Medical Sciences,
Banaras Hindu University,
Varanasi 221005, India

D. D. Pratyush
Department of Endocrinology & Metabolism,
Institute of Medical Sciences, Banaras Hindu University,
Varanasi 221005, India

Introduction

Cancer is one of the leading cause of death worldwide. It accounts for 7.9 million deaths (around 13% of all deaths) in every year and projected to continue rising, with an estimated 17 million deaths in 2030 (WHO) [1]. In the past, various anti-cancer drugs were identified and developed without focusing on a particular macromolecular target. Cancer research aimed at treatment had been the focus of many industries and academic groups since its last two decades. The treatment usually falls into one of the following categories: surgery, radiation, chemotherapy and

hormone therapy etc. [2]. The above therapies are limited by serious and life threatening side effects from toxicities to sensitive normal cells due to lack of specificity [3]. Currently, the fields of biochemistry, molecular biology, genetics and pharmacology have grown considerably in their ability to identify specific biological targets. Computational tools have recently been used to explore such targets in designing new drugs with the aim to decreased illness. Docking and molecular dynamics are the most commonly used computational tools for elucidation of cancer targets. Using the above tools, it is easier to find out the interactions and dynamics of drug and target at molecular level [4].

Actin is a cytoskeletal protein present within all eukaryotic cell types. The cell cytoskeleton is known to provide the basic infrastructure for maintaining cell morphology and functions such as adhesion, motility, exocytosis, endocytosis, and cell division. The actin cytoskeleton and its regulatory proteins are crucial for cell migration and movement in most cells. The mechanism of cell movement involves actin remodeling which actually produces necessary force for cell migration. However, it has been observed that the alteration of actin polymerization or actin remodeling plays a pivotal role in regulating the morphology and phenotypic events of cancerous cells as a result of activation of oncogenic signaling pathways, e.g., Ras and Src [5–7]. Thus, elucidation of the molecular mechanisms of actin reorganization is an important target for cancer therapeutics.

Ribonuclease displays a variety of biological functions such as degradation of RNA, control of gene expression, cell growth and differentiation, cell protection from pathogens and apoptosis etc. Besides, ribonuclease works as potential anti-tumor drugs due to their cytotoxicity and uniquely influences several functions in the tumor cells. It has been observed earlier that the various ribonucleases from different sources such as onconases (*Rana pipiens)*, bovine seminal RNase (Bovine seminal fluid), RNase T1 (*Aspergillus oryzae*), α-sarcin (*Aspergillus giganteus*), RNase P (Cultured human cells), ACTBIND (*Aspergillus niger* B1 (CMI CC 324626)) and RNase T2 (Cultured human cells) have been used for the treatment of cancer [8].

In the present study, the 3-D structure of *A.niger* ATCC 26550 RNase and human actin have been constructed using the templates ACTBIND (PDB ID: 3D3Z) [9] and crystalline profilin-beta-actin (PDB ID: 2BTF) [10], respectively. The resulting FASTA sequence was used to build the 3-D structures. Further, the molecular models of *A.niger* ATCC 26550 RNase and human actin was constructed using Modeller9v5 package for homology modeling. The model quality was assessed using PROCHECK, PROSA, ERRAT, VERIFY 3D and WHAT-IF. The overall scores were used to choose the final model. Protein-protein docking was performed between the molecular models of *A.niger* RNase and human actin. The detailed analyses of probable

inhibition as well as interaction of the models were performed with high binding affinity. The studies presented in this manuscript will be useful to design molecules that may have anticancer activity.

Materials and methods

All the calculations were performed on a workstation Hi-end server: Pentium IV 3.4 MHz, AMD Athlon 64 bits dual processor with 4 GB RAM and video graphics card. Molecular modeling tasks were performed with Modeller9v5 [11]; protein-protein docking calculations were performed with HEX. For homology modeling of the *A.niger* RNase and human actin, the crystal structures of Actinbind a T2 RNase (PDB ID: 3D3Z) and structure of crystalline profilin-beta-Actin (PDB ID: 2BTF) respectively were used as templates [9, 10]. If not otherwise stated, default settings were used during all calculations.

Sequence alignments

To find out an appropriate template structure for constructing the target model, the sequence of *A.niger* RNase (gi No. 238828131) and human actin (gi No. 313507212) were obtained from National Centre for Biotechnology Information (NCBI). With the aim to find an adequate template for homology modeling of *A.niger* RNase and human actin, sequence alignments of its amino acid sequence against Protein Data Bank [12] was performed by means of the BLAST algorithm [13, 14]. The BLASTp alignment between the selected templates *A.niger* RNase and human actin was further refined using sequence alignments in the ClustalW 2.0.12 with default parameters [15].

Molecular model building

The search using the BLASTp alignment algorithm within the PDB database showed various potential templates for molecular modeling purposes. More than 70 crystallographic structures were found to show high identity score with respect to *A.niger* RNase and human actin. Among them, ACTBIND (PDB ID: 3D3Z) and crystalline profilin-beta-actin (PDB ID: 2BTF) structures were selected as templates for *A.niger* ATCC 26550 RNase and human actin, respectively. The 3D structures of *A.niger* ATCC 26550 RNase and human actin were predicted by homology modeling on the basis of the structures of 3D3Z and 2BTF was using the program Modeller 9v5. This program is an automated approach to comparative modeling by satisfaction of spatial restraints [16–18]. The modeling procedure begins with an alignment of the sequence to be modeled (target) with relative known three-dimensional

structures (templates). 100 models were generated and among them the one having lowest root mean square deviation (RMSD) value when superposed onto the templates 3D3Z and 2BTF, was chosen for further analysis [19].

Validation of the homology model

After the construction of the model, its quality was assessed considering both geometric and energetic aspects using PROCHECK [20] ERRAT [21], WHAT-IF [22], PROSA 2003 [23] and VERIFY 3D [24, 25] for internal consistency and reliability. The Ramachandran plot computed with PROCHECK provided the residue position in particular segment based on the dihedral angles. Finally, the best-quality models were subjected to further calculations and molecular modeling studies, binding site analysis and other calculations.

The structurally conserved regions (SCRs) between reference proteins were then identified and superimposed and the *A.niger* RNase sequence was then aligned to the SCRs using the alignment module, and finally fine-tuned manually. Any one of the two segments for each SCR could be chosen to assign coordinates to *A.niger* RNase and human actin models. Then the chosen models were subjected to energy minimization in order to obtain a stable, low-energy conformation.

Protein-protein docking

A docking study was conducted to evaluate the predictive ability of the *A. niger* RNase homology model and its relevance for use in the structure-based drug design studies. In order to perform protein-protein docking between the models of *A.niger* RNase and human actin generated model

Fig. 1 (a) Sequence alignment of *A.niger* RNase with the crystallized 3D3Z. Highly conserved residues are represented in rectangular boxes. (b) Sequence alignment of human actin with the crystallized 2BTF. Highly conserved residues are represented in rectangular boxes
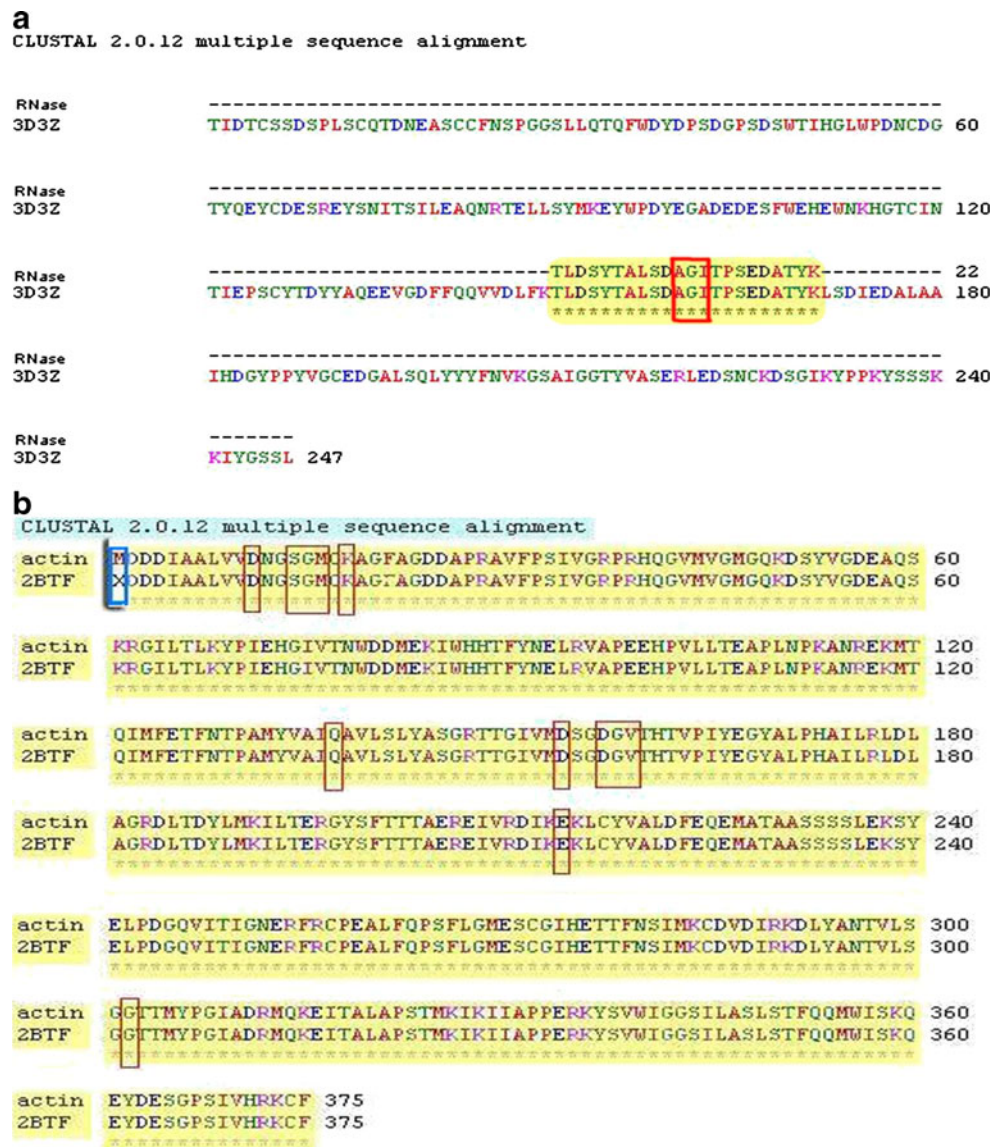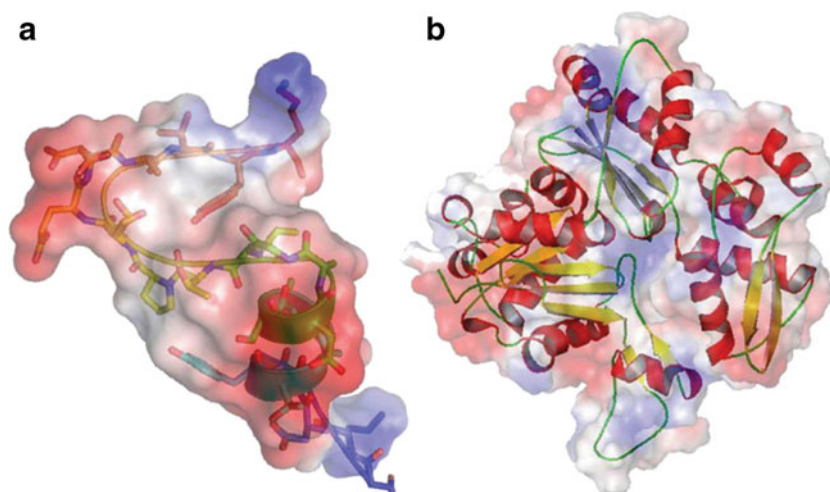
PDB's were submitted separately to online server HEX at its default parameters. HEX is an interactive protein docking and molecular superposition program; it works on FFT correlation using spherical polar coordinates and Gaussian density representation of protein shape. The computational part of the server consists of a 32-node cluster running the CentOS 5.2 operating system and using the OAR batch scheduling system (http://oar.imag.fr/). Each node consists of two quad-cores Intel Xeon 2.5 GHz CPUs, and eight of the nodes are equipped with two Nvidia Tesla C1060 GPUs. Hence, a total of 256 CPU cores and 16 GPUs are currently available on HEX server [26]. The parameters used for the docking process were Correlation type – Shape only, Calculation Device- GPU, Number of Solutions-100, FFT Mode –3D fast lite, Grid Dimension–0.6, Receptor range–180, Ligand Range–180, Twist range–360, Distance Range–40. The drug and its analogues were docked with the receptor using the above parameters. The software Pymol (http://pymol.source-forge.net/) and visual molecular dynamics (VMD) are very flexible, extensible packages for molecular visualization used to generate clear, informative and attractive representation of atomic data [27]. There have been many efforts to predict protein-protein interaction binding sites based on the analysis of the protein surface properties [28–35] analyzed the surface patches using six

parameters: solvation potential, residue interface propensity, hydrophobicity, planarity, protrusion and solvation accessible surface area. The six parameters were then combined into a global score that gave the probability of a surface patch forming protein-protein interaction.

Results and discussion

Sequence alignments

The BLASTp result of *A.niger* RNase showed 100% sequence similarity to ACTBIND (PDB ID: 3D3Z) where as in the case of human actin sequence similarity was 99% with crystalline profilin-beta-actin (PDB ID: 2BTF). Thus two structures ACTBIND (PDB ID: 3D3Z) [9] and crystalline profilin-beta-actin (PDB ID: 2BTF) were selected as templates [10] for *A.niger* RNase and human actin, respectively. The most significant step in homology modeling process is to obtain the correct sequence alignment of the target sequence with the homologues. The sequence alignment was performed using the ClustalW 2.0.12 for homology modeling as shown in Fig. 1a and b. The figure reveals that the residues involved in binding of various feedback inhibitors in template 2BTF (Asp11, Ser14, Gly15,

Table 1 PROCHECK and ERRAT

PROCHECK								ERRAT score
Ramachandran plot quality (%)				Goodness factor				
	Most favored	Additional allowed	Generously allowed	Dis-allowed	Dihedral	Covalent	Overall	
RNase	100	0.0	0.0	0.0	−0.05	0.37	0.13	84.615
3D3Z	86.5	13.0	0.0	0.5	−0.20	0.28	−0.01	97.391
Actin	94.8	4.0	0.6	0.6	−0.03	−0.22	−0.09	86.104
2BTF	84.4	13.3	1.4	0.9	−0.47	0.02	−0.63	91.111

Fig. 3 (a) Ramachandran plot of developed *A.niger* RNase, total residues are 22, in which 18 residues in most favored region (100%), end residues are 2, glycine and proline residues. (b) Ramachandran plot of developed human actin, in which 309 residues in most favored regio (94.8%), 13 residues are additionally allowed regions (4%), and reaming residues generously allowed and disallowed regions (0.6%)
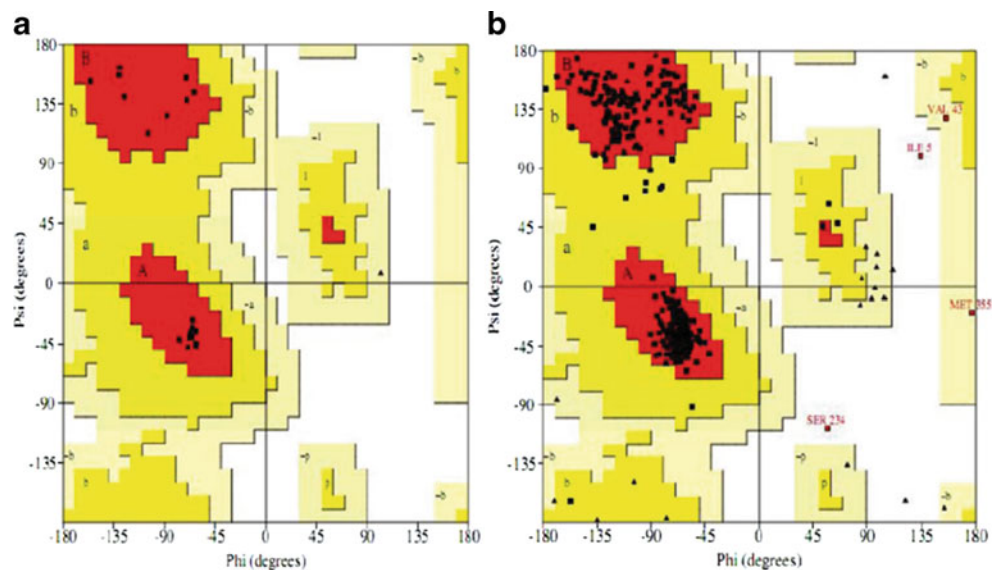
Met16, Lys18, Gln137, Asp154, Asp157, Gly158, Val159, Glu214, Gly302) were conserved in human actin. Similarly, the conserved amino acid residues involved in the binding of *A.niger* RNase and corresponding catalytic residues of template 3D3Z (Ala159, Gly160, and Ile161) have been identified.

Molecular model building

The appropriate template was chosen based on sequence similarity, residue completeness, and crystal resolution. A hundred models were generated and the model showing the least RMSD with respect to trace (Cα atoms) of the crystal structure of the template was saved for further refinement and validation (Fig. 2a and b). Further, refinement was performed in order to obtain the best conformation of the developed model of *A.niger* RNase and human actin. Among the available potential templates, crystal structure of ACTBIND (PDB code: 3D3Z, resolution 1.7Å, R-value 0.191 (obs), and R-Free-0.231) [9] was selected as the

template structure to construct molecular model of the *A. niger* RNase. The same procedure was applied for modeling the 3D structure of human actin based on the structure of crystalline profilin-beta-actin (PDB code: 2BTF, resolution 2.55Å, R-value 0.199 obs) [10].

Model quality assessment

The first validation was carried out using Ramachandran plot analysis computed with PROCHECK by checking residue-by-residue stereochemical quality of the protein structures. The analysis showed that residues of *A.niger* RNase, 3D3Z, human actin and 2BTF in the most favorable region were 100%, 86.5%, 94.8% and 84.4% and in the additional allowed region were 0%, 13%, 4% and 13.3%, respectively (Table 1 and Fig. 3a, b). The root mean squares deviation (RMSD) between the Cα atom of the template 3D3Z and the model was 0.89Å and the template 2BTF and the model was 0.92Å, which indicated high structural homology. Thus, the selected structure of *A.niger* RNase

Fig. 4 (a) Secondary structure of developed model of human actin. (b) Secondary structure of *A.niger* RNase (22 peptide Sequence). (c) Superposition of the average structure of human actin. The structures are presented as cartoon diagram. The average and the initial structure structures are colored green and magenta purple, respectively
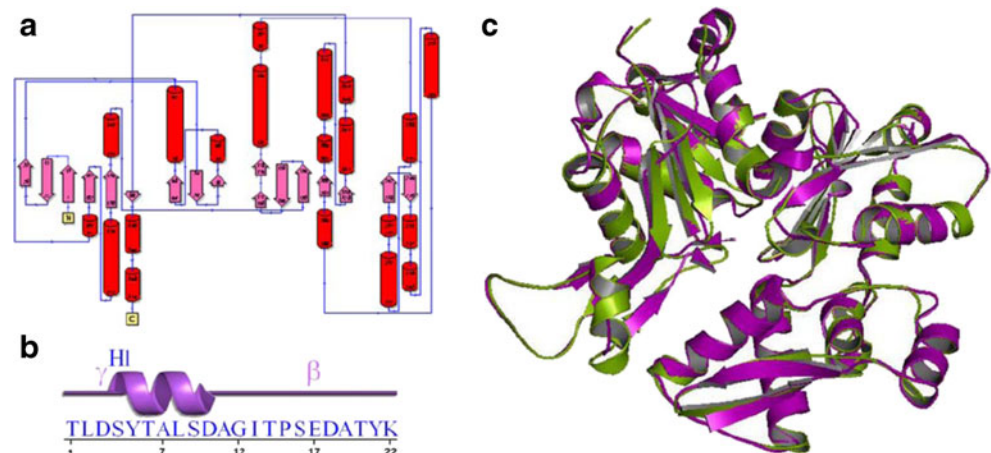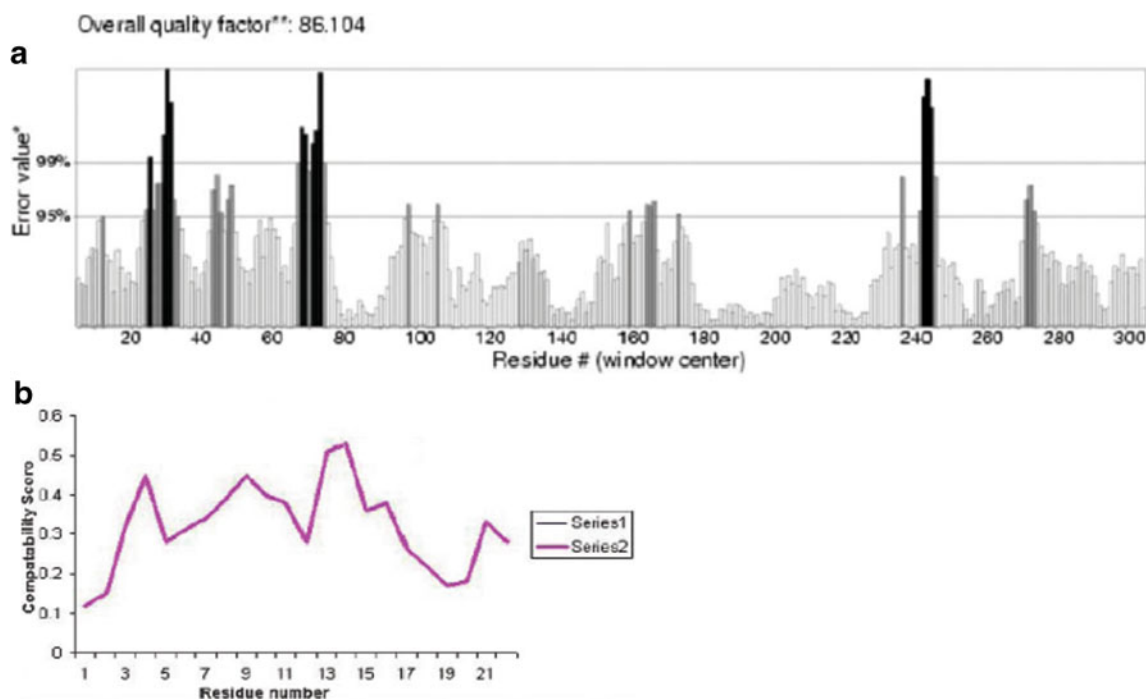
Fig. 5 (a) The ERRAT score for the human actin model is 86.104, the backbone conformation and non-bonded interactions of human actin and *A.niger* RNase homology models were all reasonable within a normal range. (b) The 3D profiles verified results of predicted human actin. The residues with positive compatibility score are reasonably folded

and human actin were found to be reasonable and reliable conformation was used for further protein docking. The selected structures were depicted in Fig. 2a and b. The total quality G-factor −0.2 indicated a good quality model (acceptable values of the G-factor in PROCHECK are between 0 and −0.5, with the best models displaying values close to zero). The PROCHECK stereochemical analysis showed no bad contacts and no bad scores for main-chain or side-chain parameters.

The detailed secondary structural investigation of the predicted human actin model with PDB sum, a secondary structure prediction server revealed 70 (18.7%) residues were in strands, 137 (36.5%) residues were in α-helices,

10 (2.7%) residues that were in 3–10 helix and 158 (42.1%) residues were in other conformations (Fig. 4a). Similarly, PDB sum secondary structure of *A.niger* RNase revealed 7 (31.8%) residues were in α-helices and 15 (68.2%) residues were in other conformations (Fig. 4b). The tertiary structure of human actin showed close resemblance to 2BTF with a backbone RMSD value 0.58Å. It was found that using the integrated sequence alignment tools and structural superposition algorithms, a target sequence can be mapped onto the modeling templates in one step. Then the initial sequence alignment can be optimized manually while the anticipated changes in the model backbone are reflected in real-time in the displayed structural superposition. By

Table 2 WHAT IF stereochemical quality evaluation

Structure	Avarage package quality[a]	Rotamer normality[b] χ_1 χ_2	Backbone conformation	Bond length[c]	Angle[d]
RNase	3.124	−0.248	−2.362	0.919	1.300
3D3Z	0.785	0.202	−0.645	0.775	0.775
Actin	1.165	−1.908	−0.480	0.923	1.255
2BTF	1.890	−5.140	−1.097	0.964	1.992

[a] The average quality of 200 highly refined X-ray structures was −0.5±0.4

[b] The behavior of the these distribution is much that a Z-score below −2 (2 standard deviations way from the average) is poor, and a Z-score of less than −3 is of concern; positive is better than average

[c] RMSD Z-score should be close to 1.0

[d] RMSD Z-score, more common values are around 1.55

Fig. 6 (a) PROSA-web Z-scores of all protein chains in PDB determined by X-ray crystallography (light blue) and NMR spectroscopy (dark blue) with respect to their length. The Z-score of human actin was present in the range represented in large black dot. (b) Energy plot for the predicted human actin
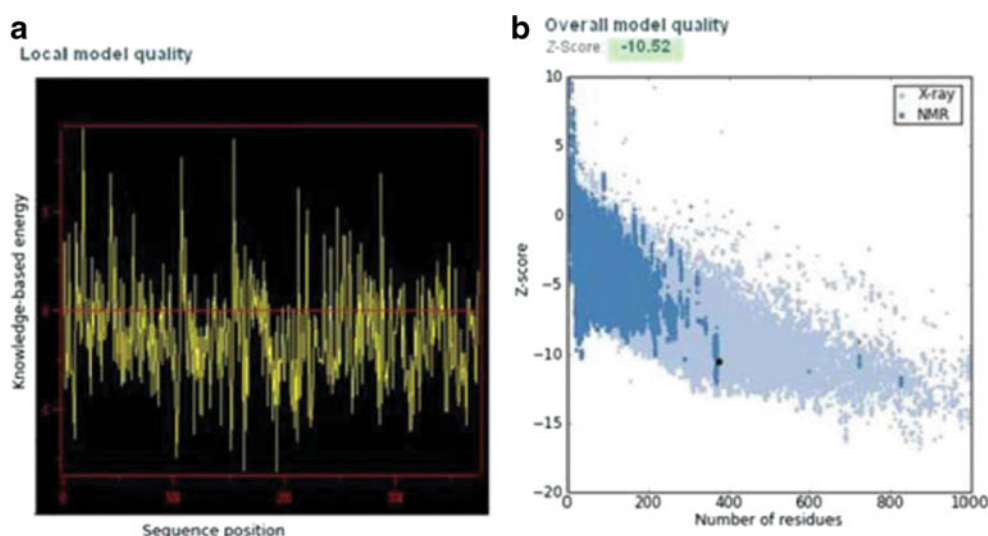
applying structural superposition and RMSD evaluations, our model appears very similar to the experimental one. The superposition of the average structure of the human actin with the initial model did not show major structure conformational changes in comparison to the initial model, which in turn is consistent with the relatively low RMSD values (Fig. 4c). The overall low RMSD values for backbone superposition reflect the high structural conservation of this complex through evolution, making it a good system for homology modeling.

ERRAT is a so-called "overall quality factor" for non-bonded atomic interactions, and higher scores mean better quality [21]. The normally accepted range for a high quality model is >50 [36]. In the current case, the ERRAT score for the human actin and A.niger RNase models were 86.104 and 84.615, respectively and fit well within the range of a high quality model as shown in Fig. 5a. Whereas the ERRAT score for the templates 2BTF and 3D3Z were 91.111 and 97.391, respectively (Table 1). Thus, the above analysis suggests that the backbone conformation and non-bonded interactions of human actin and A.niger RNase homology models were all reasonable within a normal range.

The final evaluation of the built human actin structure was checked by VERIFY 3D [24, 25]. The VERIFY 3D analysis indicated a reasonably good sequence-to-structure agreement because none of the amino acids had a negative score (average score=0.36) as shown in Fig. 5b. It is to be noted that compatibility scores above zero correspond to acceptable side chain environment.

WHAT-IF is used to check the normality of the local environment of amino acids [26]. For the WHAT-IF evaluation, the quality of the distribution of atom types was determined around amino fragments. For a reliable structure, the WHAT-IF packing scores should be above −5.0 [22]. In this case, none of the scores for each residue in the homology model is lower than −5.0 as depicted in Table 2. Therefore, the WHAT-IF evaluation also showed that the homology mode structure was very reasonable. Analysis of the energy minimized human actin model with WHAT-IF web interface revealed that RMSD Z-Scores for bond angles and bond lengths were all close to 1.0 and also within the limits of templates. The interaction energy per residue was also calculated by the PROSA2003 program [23]. In this analysis,

Table 3 Protein-Protein Docking Results from HEX server

Receptor protein	Lead protein	Cluster	Solution	RMSD from reference structure(Å)	Etotal (kcal/mol)	Docked energy (kcal/mol)
Human actin	A.niger RNase	1	1	0.614	−1.873	−2.333
		2	3	0.489	−1.822	−6.442
		3	4	0.698	−1.762	−8.117
		4	5	0.232	−1.739	−6.363
		5	6	0.984	−1.736	−5.204
		6	7	0.389	−1.696	−8.332
		7	10	0.704	−1.671	−8.496
		8	11	0.654	−1.653	−9.930
		9	14	0.562	−1.652	−2.366
		10	16	0.617	−1.608	−8.369

Fig. 7 Interacting amino acid residues on the RNase actin and RNase proteins. The interacting region of RNase is represented in blue color sticks and the regions of human actin are shown by cartoon in which catalytic residues in spherical form. The 3D structure of the RNase and actin complex was predicted by Protein docking using HEX software
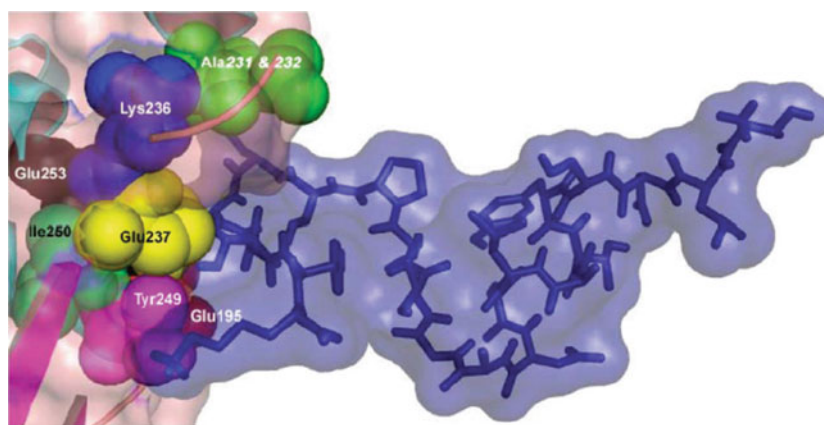
the interaction energy of each residue with the remainder of a protein was computed to judge whether it fulfills certain energy criteria or not. The PROSA Z-Score indicates overall model quality. Global analysis of the model human actin with PROSA showed a Z-Score of −10.52, indicating no significant deviation from typical native structures of similar size as the template when compared with Z- Scores of −10.4 for 2BTF template. Figure 6a displays the PROSA2003 energy profiles calculated for the human actin model along with the templates. The energy profile of the human actin homology model was consistent with a reliable conformation based on its similarity to that of the template 2BTF. Evaluation of the energy minimized model of human actin with PROSA-web revealed that the Z-score value was −10.52 (Fig. 6b) in the range of native conformations of the crystal structures. PROSA-web analysis (Fig. 2a) had showed that overall the residue energy of the human actin model was largely negative except for some peaks in the middle region.

Protein-protein docking

In order to understand the inhibition mechanism of *A.niger* RNase on human actin, primary docking calculations were performed with HEX. The goal of the initial stage of docking is to generate as many near-native complex structures (hits) as possible. The generated PDB file was analyzed for their binding conformations. Analysis was based on Etotal or free energy of binding, lowest docked energy, and calculated RMSD values. For each approach, the number of hits, the RMSD value of the best hit (with the lowest RMSD) based on shape complementarity are listed in Table 3. The results obtained from protein–protein docking algorithms were satisfactory. The total clusters of docking conformations, with the top 30 docked molecules showed negative binding energies. Cluster 1 shows the energy and RMSD values to be −1.873 kcal mol^{-1} and 0.614 Å, respectively. Among all docking clusters, rank 10, i.e., solution 16 gave the best predicted binding free energy of −1.608 kcal mol^{-1} with RMSD value of 0.617 Å. *A.niger* RNase docking revealed

that the amino acids Ser16, Glu17, Asp18, Ala19, Thr20, Tyr21, and Lys22 played vital role to bind the Glu195, Ala231& 232, Lys236, Glu237, Tyr249, Ile250 and Glu253 of human actin (Fig. 7). The ball and stick model of *A.niger* RNase is presented in Fig. 8a. The docking of *A.niger* RNase and human actin is shown in Fig. 8b. Our in-silico experiments demonstrate that *A.niger* RNase binds human actin,
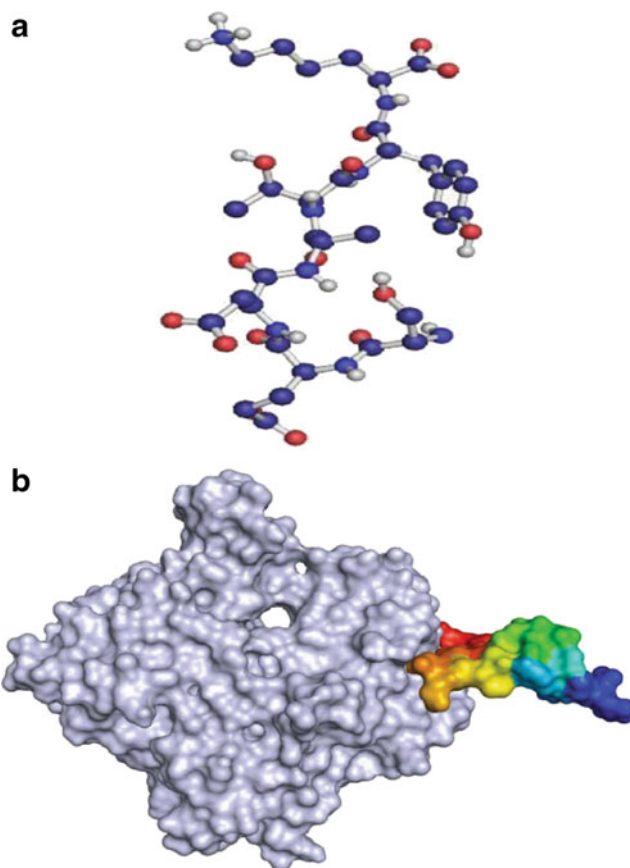
Fig. 8 (**a**) *A.niger* RNase represented in ball and stick form, in which carbon-density blue, hydrogen-white, nitrogen-blue, oxygen-red and sulfur-orange color. (**b**)Protein-protein docking interaction of *A.niger* RNase represented in rainbow and human actin in vacuum electrostatic light blue white color

and also is itself inhibits its function and thus may act as a drug. Cumulatively, the evidence leads us to propose that cell-surface actin may be a target for RNase in cancer cells. Therefore, A.niger RNase may compete with angiogenin for cell-surface actin and, in this manner, blocks the formation of the actin-angiogenin complex required for cancer cell organization and angiogenesis in developing neoplastic tissue.

Conclusions

In this study, we have developed very high accurate 3D models of human actin protein and A.niger RNase by using the templates ACTBIND (PDB ID: 3D3Z) and crystalline profilin-beta-actin (PDB ID: 2BTF), respectively in Modeller9v5. This model has been qualified using several validation methods, including PROCHECK, ERRAT, WHAT-IF, PROSA2003 and VERIFY-3D. All evidences suggest that the geometric quality of the backbone conformation, the residue interaction, the residue contact and the energy profile of the structure is well within the limits established for reliable structures. The protein-protein docking study has been elucidated for the purpose of finding anticancerous property of A.niger RNase. The interaction energy of docking between the A.niger RNase and human actin was calculated and analyzed using the HEX server. The efficient binding of A.niger RNase and human actin revealed that the proteins could form hydrogen bond networks involving active amino acid residues. Several amino acid residues including Ser16, Glu17, Asp18, Ala19, Thr20, Tyr21, and Lys22 were identified to exclusively contributive to the binding of A. niger RNase to Glu195, Ala231 & 232, Lys236, Glu237, Tyr249, Ile250, and Glu253 of human actin. The generated homology model is expected to be useful for the structure-based drug design against cancer.

References

1. World Health Organization (http://www.who.int/cancer/en/)
2. http://www.medicalnewstoday.com/info/cancer-oncology/
3. Allen TM (2002) Ligand targated therapeutics in anticancer therapy. Nat Rev Cancer 2:750–763
4. Peitsch MC, Schwede T, Guex N (2000) Automated protein modeling the proteome in 3D. Pharmacogenomics 1:257–626
5. Pollard TD, Borisy GG (2003) Cellular motility driven by assembly and disassembly of actin filaments. Cell 112:453–465
6. Rao J, Li N (2004) Microfilament actin remolding as a potent target for cancer drug development. Curr Cancer Drug Targets 4:345–354
7. Gundampati RK, Sharma A, Kumar M, Debnath M (2011) Extracellular poly (A) specific ribonuclease from Aspergillus niger ATCC 26550: Purification, biochemical, and spectroscopic studies. Process Biochem 46:135–141
8. Roiz L, Smirnoff P, Bar-Eli M, Betly S, Shoseyov O (2006) Actibind, an Actin-binding Fungal T₂-RNase with Antiangiogenic and Anticarcinogenic Characteristics. Cancer 106:2295–2308
9. Gonzalez A, Almog O, Crystal structure of Actibind a T2 RNase, (http://www.pdb.org/pdb/explore/explore.do?structureId=3D3Z)
10. Schutt CE, Myslik JC, Rozycki MD, Goonesekere NC, Lindberg U (1993) The Structure Of Crystalline Profilin-Beta-Actin. Nature 365:810–816
11. http://www.salilab.org/modeller/9v5
12. Berman HM, Wetbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
14. 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 data base search programs. Nucleic Acids Res 50:3389–3402
15. 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
16. Sali A, Overington JP (1994) Derivation of rules for comparative protein modeling from a database of protein structure alignments. Protein Sci 3:1582–1596
17. Sali A (1995) Modeling mutations and homologous proteins. Curr Opin Biotechnol 6:437–451
18. Sali A, Potterton L, Yuan F, Van Vlijmen H, Karplus M (1995) Evaluation of comparative protein modeling by MODELLER. Proteins 23:318–326
19. Guex N, Peitsch MC (1997) SWISS MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723
20. 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
21. Colovos C, Yeates TO (1993) Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519
22. Vriend G (1990) WHATIF: a molecular modeling and drug design program. J Mol Graph 8:52–56
23. Sippl MJ (1993) Recognition of errors in three-dimensional structures of Proteins. Proteins 17:355–362
24. Bowie JU, Lüthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253:164–170
25. Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356:83–85
26. HEX server (http://www.loria.fr/~ritchied/hex_server/)
27. Humphrey W, Dalke A, Schulten K (1996) VMD-Visual Molecular Dynamics. J Mol Graph 14:33–38
28. Lichtarge O, Bourne HR, Cohen FE (1996) An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol 257:342–358

29. Jones S, Thronton JM (1997) Analysis of protein-protein interaction sites using surface patches. J Mol Biol 272:121–132
30. Jones S, Thromton JM (1997) Prediction of protein protein interaction sites using patches analysis. J Mol Biol 272:133–143
31. Neuvirth H, Raz R, Schreiber G (2004) Promate: A structure based prediction program to indentify the location of protein-protein binding stes. J Mol Biol 338:181–199
32. Bradford J, Westhead D (2005) Improved prediction of protein-protein binding sites using a support vector machines approach. Bioinformatics 21:1487–1494
33. Zhou H, Shan Y (2001) Prediction of protein interaction sites from sequence profile and residue neighbor list. Proteins 44:336–343
34. Espadaler J, Romero-Isart O, Jackson RM, Oliva B (2005) Prediction of protein-protein interactions using distant conservation of sequence patterns and structure relationships. Bioinformatics 21:3360–3368
35. Aytuna AS, Gursoy A, Keskin O (2005) Prediction of protein-protein interactions by combining structure and sequence conservation in protein interfaces. Bioinformatics 21:2850–2855
36. Laskowski RA, Watson JD, Thornton JM (2005) ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res 33:89–93

Ab initio parameterization of YFF1, a universal force field for drug-design applications

Olexandr Ya Yakovenko · Yvonne Y. Li ·
Alexander A. Oliferenko · Ganna M. Vashchenko ·
Volodymyr G. Bdzhola · Steven J. M. Jones

Abstract The YFF1 is a new universal molecular mechanic force field designed for drug discovery purposes. The electrostatic part of YFF1 has already been parameterized to reproduce *ab initio* calculated dipole and quadrupole moments. Now we report a parameterization of the van der Waals interactions (vdW) for the same atom types that were previously defined. The 6–12 Lennard-Jones potential terms were parameterized against homodimerization energies calculated at the MP2/6-31 G* level of theory. The Boys-Bernardi counterpoise correction was employed to account for the basis-set superposition error. As a source of structural information we used about 2,400 neutral compounds from the ZINC2007 database. About 6,600 homodimeric configurations were generated from this dataset. A special "closure" procedure was designed to accelerate the parameters fitting. As a result, dimerization energies of small organic compounds are reproduced with an average unsigned error of 1.1 kcal mol^{-1}. Although the primary goal of this work was to parameterize nonbonded interactions, bonded parameters were also derived, by fitting to PM6 semiempirically optimized geometries of approximately 20,000 compounds.

Keywords Force fields · Parameterization · Quantum chemical calculations

O. Yakovenko (✉) · Y. Y. Li · S. J. M. Jones
Genome Sciences Centre, BC Cancer Agency,
Suite 100 570 West 7th Avenue,
Vancouver, British Columbia, Canada
e-mail: ayakovenko@bcgsc.ca

O. Yakovenko
e-mail: yakovenko.alexander@gmail.com

A. A. Oliferenko
Department of Chemistry, University of Florida,
Gainesville, FL 32611, USA

V. G. Bdzhola
Institute of Molecular Biology and Genetics of National Academy
of Sciences of Ukraine,
150 Zabolotny str,
Kyiv 03143, Ukraine

G. M. Vashchenko
Department of Biochemistry and Molecular Biology,
University of British Columbia,
2350 Health Sciences Mall,
Vancouver, British Columbia, Canada

Introduction

The common approach in biomolecular modeling is using molecular mechanical force fields (FF): sets of predefined functions and parameters derived from experimental data or high level *ab initio* calculations. There are a lot of FFs that differs in parameterization sources, interaction modeling functions and consequently field of applicability. For instance, the MM2 parameterized to reproduce conformational energies of hydrocarbons was then extended (MM3, MM4) to work with many other organic molecules [1–4], OPLS [5, 6] was fitted against enthalpies of vaporization. GROMOS [7] is designed for modeling biomolecules in aqueous solutions. DRF90 [8] attempts to introduce polarization in the form of effective polarizability tensors, while AMBER [9] does the same with a point charge flow approach, and CHARMM with the Drude oscillator model [10]. However, it is difficult to guarantee a reasonable accuracy of a FF for problems beyond its

applicability area, which is primarily defined by parameterization.

The common technique of FF parameterization can be traced back to the early 1980 with works of Allinger [11], Weiner [12] and others contributors (see [13] for details). The main idea had been to fit parameters for small datasets of typical chemical compounds such as amines, acids, ethers, hydrocarbons, amino acids, nucleic acids, etc., and thus to come up with a limited set of chemical building blocks. A drawback of that early approach was a dependence of accuracy of calculated energy terms on the similarity of treated molecules and building blocks used in parameterization. Over the recent decade, such force fields as AMBER [14], OPLS [5], CHARMM [15] have evolved as 'general purpose' FFs that are less sensitive to such a similarity. The main reasons of this trend is the wide spread of virtual screening applications which raised a strong practical interest to universal FFs as tools for unbiased modeling of complexes of typical biological molecules (which is a relatively tractable problem) with broad range of various chemical substances (which is a challenge) [16]. Unfortunately, modifying of the present fragment parameterized FFs requires enormous amount of human efforts: to maintain backward compatibility one has to process manually numerous small sets of typical molecules [17]. The most popular fragment parameterized general FFs at the moment are GAFF [18], CHARMM [17], CFF [19], and MMFF [20].

Our motivation for a new FF development has been a special interest in potential energies of intermolecular interactions in molecular docking applications [21] as well as a lack of intentionally parameterized universal FFs. As universal we will refer to a FF which accuracy is almost independent of similarity between an analyzed molecule and all those used in parameterization. As YFF1 inherits accuracy and universality from the previously developed Kirchhoff charge model (KCM) [22], here we just add van-der-Waals (vdW) atom-specific and bonded parameters. As a result, YFF1 can be thought as free of any structural or functional group classification, and all atomic, bond length, bond angles, improper and torsion angle parameters are atom-wise with some electronic and topological features taken into account (similarly to how it was done in [23]).

Although binding free energy cannot be directly derived from molecular mechanics potential energy, its relative value can be approximately calculated by averaging of the Coulomb and vdW interaction energies with linear interaction energy (LIE) method [24, 25]. As averaging can be taken either over the time or over the ensemble, a deeper search in molecular docking configurations scanning routine somewhat resembles the LIE approach. Thus we assume a future growth of interest to fast but accurate potential energy evaluations in virtual screening. As for now, intermolecular energy is a common feature of scoring functions of all most popular molecular docking suites: DOCK [26], AutoDock [27], FlexX [28], Surflex [29], GOLD [30], ICM [31] etc. Thus YFF1 is fitted to reproduce gas-phase dimerization energies of drug-like compounds calculated at the MP2/6-31 G* theory level. An exhaustive study of water homodimerization energy [32] assumes that the 6-31 G basis set overestimates potential energy by about 15% (~0.7 kcal mol^{-1}) comparing with the 6-311 G and the aug-cc-pVTZ. On the other hand, it is several times faster and almost free of linear dependencies which often breaks QM calculations. YFF1 needs some additional tuning for using in condense phase studies, as it is known that about 30% of potential energy calculated in the gas-phase is missing compared to that observed in the condensed phase [16], [33], [34]. Some works [33] assume that diffuse orbital overlaps beyond the vdW envelope are responsible for this, but others suggest the presence of hydrogen bonds with better electrons localization [34]. Still there is no conventional scaling of gas-phase potential energies for drug-design purposes, as the amount of hydrogen bonds at the protein-ligand interface is smaller than in aqueous media.

Materials and methods

Electrostatic interactions

YFF1 uses the same atom types as in the previously published Kirchgoff charge model (KCM) [22], with details given in Table S1 of Supporting information. It inherits all parameters, and hence the accuracy of the Coulomb interactions from the previous work, where they were fitted to reproduce dipole and quadruple electric moments calculated at the HF/6 G-311** theory level. However, the KCM algorithm was modified in order to adapt it to handle larger molecules. The fast KCM (FKCM), which is now of a linear complexity with respect to the number of atoms in a molecule, is made up with two improvements. The first one is the so-called sparse matrix representation of the internal bond hardness operator, which maps effective atomic electronegativities onto point charges. Performing only those multiplications that are really necessary dramatically increases the efficiency of the KCM in dealing with large molecules (represented by highly sparse adjacency matrices). The second one is the Sherman-Morrison formula [35] as applied to analytical calculations of virtual chemical hardness. The Sherman-Morrison formula computes the inverse of the sum of an invertible matrix and the scaled dyadic product of two vectors. As the hardness matrix is perturbed by a known

dyadic product of two vectors, it is thus a linear function of the scaling constant. So it is possible to calculate analytically such a perturbation of a given unperturbed matrix that satisfies the condition that its inverse perturbed product with a given electronegativity vector equal to the formal charge of a counter-ion. Although it allows analytical expression for perturbations caused by only one virtual bond, multiple perturbations can be calculated in a few iterations of gradient optimization methods using once perturbed values as an initial guess. More detailed description of FKCM is given in Appendix A.

Van der Waals interactions

The vdW interactions were parameterized to fit homodimerization energies calculated at the MP2/6-31 G* level of theory using the PC GAMESS/Firefly quantum chemical package [36, 37]. Energy of homodimerization is the work required to bring two identical molecules into close proximity (homodimeric configuration) from infinite separation in vacuum. For each homodimeric configuration the Boys-Bernardi counterpoise BSSE correction [38] was calculated. Homodimeric configurations were generated by rigid-body molecular docking of monomers and selected as configurations of the lowest potential energy calculated by the current YFF1 parameterization. Monomers were selected from 19,374 neutral compounds retrieved from the ZINC2007 database [39], optimized with MOPAC2009 software using the PM6 semiempirical method [40]. The minimization was performed with the PRECISE keyword. For each optimized monomer, an energy gradient was calculated with PC GAMESS/Firefly at the MP2/6-31 G* level, and only those were accepted, which had the gradient norm less than 1 Hartree/Bohr. As a global optimization routine for the parameter fit, we used the Nelder-Mead simplex method [41].

Bonded interactions fit (see Appendix B for details)

The hard bonded terms were parameterized on the basis of 19,374 monomers which optimized and distorted energies were calculated with the PM6 method. The soft bonded terms, dihedral angles were entirely borrowed from the MMFF94 force field [20]. Alternatively, one can consider borrowing the bonded parameters from other force fields, as used to be done elsewhere [8].

Dimeric dataset generation

A rigid body molecular docking was used to generate a training set suitable for fitting van der Waals parameters. For each accepted monomer the coordinates (in the center of mass frame) were used for making homodimers with the second monomer placed by rotational and translational changes of the first monomer coordinates:

$$r' = R_x(\alpha_1)R_y(\beta_1)r + R_x(\alpha_2)R_y(\beta_2)(l \cdot \vec{t}), \tag{1}$$

where r' is the vector of atomic coordinates of second monomer, r is the vector of atomic coordinates of the first, static monomer, R_x is the matrix that rotates coordinate vectors around the (1,0,0) vector by angle α, R_y is the matrix that rotates coordinate vectors around the (0,1,0) vector by angle β, α_1, β_1, α_2, β_2 are the rotation angles, l is the distance between monomers centers of mass, and t is the (0,0,1) vector. This five-dimensional space was sampled in 12 points along each axis to form a tractable for exhaustive search multitude of 62,208 configurations per monomer, formed by exhaustive rigid body rotations (with the step of $\pi/6$) and center-of-mass translations (with 1 Å steps) from 0 to 12 Å.

Results and discussion

The force field

YFF1 is aimed to work in the internal cycles of molecular docking algorithms and thus to model atomic pair-wise interactions in a very accurate manner. For these purposes, the nonbonded terms are of primary interest, while the bonded one (except soft bonded) would be of less importance, as docking algorithms usually perform searches in internal coordinates. Our choice for the nonbonded interactions is the well-known 6–12 Lennard-Jones potential as well as the classical nonpolarizable expression for the Coulomb interactions. For bonded interactions the robust harmonic potential functions were accepted. Similar energy terms are used in most of modern force fields and they normally provide reasonable structures in a majority of practical cases. The exact formulation of the YFF1 is as follows:

$$
\begin{aligned}
U &= U_{el} + U_{vdw} + U_{imprs} + U_{tors} + U_{angle} + U_{bond} \\
U_{bond} &= K_b(r - r_0)^2 \\
U_{angle} &= K_a(\alpha - \alpha_{0l})^2 \\
U_{tors} &= K_{\varphi 1}\cos\varphi - K_{\varphi 2}\cos 2\varphi + K_{\varphi 3}\cos 3\varphi \\
U_{imprs} &= K_i\varphi^2 \\
U_{el} &= K\frac{q_iq_j}{r^2} \\
U_{vdw} &= \frac{A_iA_j}{r^{12}} - \frac{B_iB_j}{r^6}
\end{aligned}
\tag{2}
$$

In this study we are focused on the parameterization of nonbonded interactions and, therefore, refer to a FF parameterization as to a fit of Lennard-Jones parameters, unless otherwise stated. So we need to parameterize in (2)

one repulsive constant A and one attractive constant B per atom type. However, in practice, nonbonded energies are not evaluated strictly as it described there. For the sake of calculation speedup, explicit square matrices of A and B constants are formed and then pair-wise energies are evaluated in the following way:

$$U_{vdw} = \left(A_{ij}^* r_{ij}^{-6} - B_{ij}^* \right) r_{ij}^{-6}, \qquad (3)$$

where the asterisk means a pre-calculated pair-wise value. Thus one can consider a FF parameterization as finding values of all elements in the A^* and B^* matrices (i.e., $N(N+1)$ parameters, where N is the number of the FF atom types). If one obtains parameters in the form of (2), relatively 'rigid' but easily extendable FF can be produced (i.e., it is quite easy to add new atom types into an FF). Alternatively, it is possible to parameterize FF more accurately by fitting all $N(N+1)$ parameters in the form of (3), but the ability to extend such FF without total re-parameterization would be lost. Moreover, fitting all A^* and B^* is coupled with a dramatic increase in the number of parameters and thus with a significant reduction of the fitting routine efficiency. This is why the current parameterization chooses an intermediate way. The kernel of the FF is built according to (2), i.e., appropriate A and B constant are sought for every atom. In addition, for the B^* matrix elements, which correspond to hydrogen donor-acceptors and to some most frequently occurred atom pairs, particular B_{ij} values were fitted (totally 74 additional B_{ij} values). This choice is supported by the fact that the B constants are more diverse. In addition, this approach allows one to remove to some extent the bias caused by an uneven distribution of atom types.

Fitting strategy

Obviously, empirical FFs, irrespective of parameterization details, cannot arbitrarily accurately approximate QM energies. Thus the goal is to find a natural limit of the accuracy with the intent to minimize amount of expensive quantum chemical calculations. In other words, it is a search for the smallest set of homodimeric structures that would be "closed" (mathematically speaking) with respect to the parameterization routine. This is such a minimal set of molecular configurations for which the best fitted parameters again lead to the members of the same set in search for global energy minima. The search for a "closed" configurations set and for the corresponding parameters is realized as an iterative procedure.

Let assume a training set at an ith iteration. For this training set the best parameters are fitted to interpolate a QM potential energy. Then a search for global energy

minimum with the newly acquired parameters reveals some amount of new configurations that were not considered previously. If this amount is reasonably big, the algorithm has to do the $(i+1)$th fitting iteration. Otherwise, this set of parameters generates a set of configurations that is "closed" with respect to parameterization and the parameterization routine stops. The parameters obtained are considered as the best for this closed set and the corresponding accuracy will be the limit of accuracy for a given FF. The only condition one needs to satisfy is a reasonable amount of configurations considered with an exhaustive search while looking for a global minimum. On the one hand, it has to be large enough to represent a vast variety of possible dimeric configurations, but on the other hand it has to be reasonably small to scan it exhaustively with modern computers (otherwise the set can be closed by luck). A rigid-body molecular docking procedure was used to scan the configuration space spanned by approximately $6.2 \cdot 10^4$ points per dimer. Parameters borrowed from GROMOS96 [7] were used as an initial guess and initial training set was an empty set. The Nelder-Mead simplex method was used as a fitting routine. This method is known for producing good results in problems of global minimization and it does not require a gradient of the object function. Although simplex is a rather slow method, it is not a problem, because the greatest computational complexity of the protocol is *ab initio* calculations of dimerization energy.

The main task to be solved at each iteration is to find the best parameters for the interpolation of quantum chemical energies. To obtain new parameters three consecutive simplex runs were applied. The first simplex run takes the previous FF's kernel for an updated set of dimeric configurations and optimizing the additional B_{ij} parameters only. The second simplex run updates the kernel itself by changing atom's B and A constants while keeping the B_{ij} parameters frozen. New kernel parameters and additional corrections are then optimized simultaneously at the third simplex run. The lowest final point of a previous simplex in series is then used as an initial condition for the next one. The following objective function was used for simplex optimization:

$$s = \sum_i \begin{cases} \left| U_i^{QM} - U_i^{FF} \right| > 0.24 & \left(\ln \left[4.184^2 \cdot \left(U_i^{QM} - U_i^{FF} \right)^2 \right] \right)^2 \\ \left| U_i^{QM} - U_i^{FF} \right| \le 0.24 & 0 \end{cases}$$

$$(4)$$

where U^{QM} is the quantum chemically calculated dimerization energy, U^{FF} is the force field calculated dimerization energy, and index i runs over all configurations in the dataset. This function avoids biases of small fraction of configurations interpolated with greater errors. On the other

hand, it is still sensitive to the absolute value of error. Interpolation errors below 0.24 kcal mol^{-1}·were ignored to avoid a negative score.

The only serious problem that such strategy would run into is the difference in the description of molecular systems on the molecular mechanical and quantum mechanical levels. As an example, two molecules that are located too close at quantum chemical level might be treated as a transition state of a chemical reaction, but for molecular mechanics it is just an elastic collision of two molecules. To handle this discrepancy, all dimeric configurations were rejected if monomers were located too close and the FF interpolation energies differed significantly (about two orders of magnitude) from those calculated *ab initio*. After deleting such "outliers" the fit was repeated. Only one deleting procedure was applied at each iteration, and only those configurations were deleted that had been added during the current iteration. As the procedure converges, the amount of such colliding dimers vanishes.

Dataset generation

This sort of iterative parameterization protocol requires high quality homodimerization energies. They cannot be good unless monomers used for rigid docking are themselves properly minimized. The computational complexity of energy minimization at high levels of *ab initio* theory prohibits optimization of high-throughput generated complexes. Thus our homodimeric dataset could suffer of two problems: (i) insufficient geometry optimization of monomers, and (ii) neglecting geometry perturbations on the complex formation. Since the second problem is inevitable because of the prohibitive complexity of quantum chemical calculations, there is no special reason to spend resources to generate exceedingly accurate monomeric geometries. The latter should be "not too bad" to avoid significant errors on the complex formation. So we minimized geometries of *ca*. 20,000 neutral compounds retrieved from the ZINC2007 database using the semi-empirical method PM6 implemented in MOPAC2009 software. As for any nontrivial compound several local minimums exist and gradient optimization reveals only closest local one, the training set was generated using monomers at local, not necessary global, minima. Then energy gradients were calculated at the MP2/6-31 G* level of theory with PC GAMESS/Firefly software. For the parameterization purposes we have selected 2,396 monomers with the gradient norm less than 1.0 Hartree/Bohr. The ZINC2007 database is a reliable source of typical compounds used in drug design and thus typically passed through virtual screening and molecular docking software which YFF1 is designed for.

Table 1 Iterative fit of parameters

Iteration number	1	2	3	4	5
Total number of configurations	2,396	4,308	5,611	6,354	6,606
Average fit error, kcal/mol/mol	1.08	1.30	1.17	1.18	1.18
Average fit square error, kcal^2/mol^2	3.11	11.44	4.84	5.22	5.22
Number of added configurations	2,396	2,092	1,346	753	252
Number of deleted configurations	0	180	43	10	0

Iterative fit results

The iterative fit results are given in Table 1. At the initial iteration dimer formation was constrained so distances between two closest atoms of different molecules were greater than 3Å plus the sum of their van der Waals radii. With this constraint, generated configurations were *a priory* "not too bad" in terms of dimerization energy (i.e., without overlaps) and thus positive features were inherited from GROMOS96. Due to the absence of restrictions at the second iteration the algorithm eventually meets "bad" configurations. These usually appear due to atoms overlaps. Thus approximately 200 overlapping configurations were discarded at the second iteration. All subsequent iterations were done to find a compromise between "bad" overlapping configurations and the increasing strength of interactions on atoms coming closer to each other. The convergence criterion is the number of newly generated unique configurations at the current iteration. It decreases quickly rendering a super-linear convergence. After the fifth iteration new configurations were only a small fraction (92 unique configurations) of the entire monomers set. The changes in parameters are made accordingly to the number of new dimeric structures and were almost frozen.

The final YFF1 parameters are listed in Table S1 of the Supporting information along with the Coulomb parameters and atom type descriptions. Figure 1 illustrates accuracy of homodimeric QM energy interpolation with YFF1. The mean unsigned error of the fit is 1.10 kcal mol^{-1} (the average square error 2.94 kcal^2 mol^{-2}), with the deletion of 0.5 % of the strongest outliers (34 configurations) taken into account. These configurations are due to the discrepancy between the quantum chemical and force field descriptions and are scored on average by chance at the iterations where they appeared and were not deleted.

Parameters

As YFF1 tends to have universally fitted parameters, their errors do not depend on the chemical nature of a molecule.
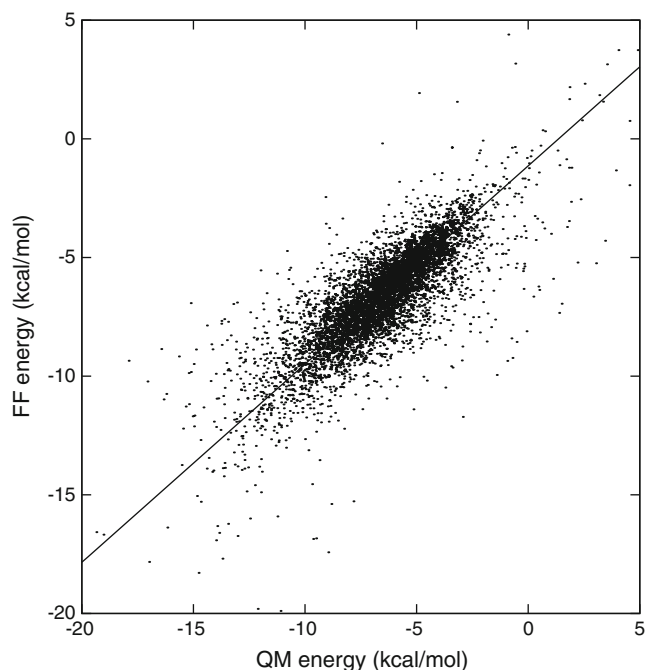
Fig. 1 Force field vs. quantum chemical dimerization energies (kcal mol^{-1})·of the closed training set with 0.5 % of strongest outliers deleted

There is no correlation between the value of the YFF1 dimerization energy error and either the van der Waals energy (in the Lennard-Jones form) or the Coulomb interaction energy, which stabilize the dimers. Moreover, there is no correlation between the error and the presence of the biologically important π-π stacking interactions. The stacking component cannot be directly extracted from quantum chemical energies. However, the information on a significant number of closely spaced molecular orbitals can be deduced by the comparison of BSSE corrections and energies of single monomers. There is also no correlation between the YFF1 dimerization energy error and the value of BSSE correction compared to the twice as the single monomer energy (both monomers' electronic energies are the same because dimeric configurations are made *via* a
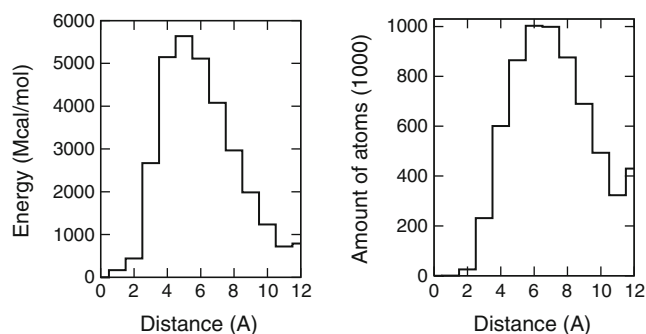
Fig. 2 Distance distribution (with a 1.0 Å sampling) of the interaction energy and the amount of interacting atoms pairs in the closed training set

rigid body docking). BSSE corrections itself are essential for the evaluation of dimerization energy, because the average difference between the sum of BSSE energy corrections for both monomers and the twice as monomer energy is −9.14 kcal mol^{-1} for the closed training set (~137 % of the average dimerization energy). The dimerization energies are usually overestimated unless BSSE corrections are subtracted.

The distances between atoms used for the parameterization densely cover the range from 2.0 to 12.0 Å (see Fig. 2). The majority (approximately $4 \cdot 10^6$) of the interacting atom pairs are separated by 5.0 - 8.0 Å, although the majority of dimerization energies is sourced from distances of 4.0 − 6.0 Å. The dense distribution of both energy and atoms along with the uniform errors distribution suggests the YFF1 is able to reliably evaluate pair interactions in the range of 3.0 − 11.0 Å.

Cation-π interactions

The incorporating of cation-π interactions is somewhat an *ad hoc* procedure to the parameterization; it was added separately after a high-throughput fitting. Cation-π interactions can be naturally incorporated into YFF1 by varying the repulsive exponent A_{ij} of the charged nitrogen atom (the only allowed cationic groups in the YFF1 are R-NH_3^+, R-$CN_2H_4^+$ and R=NH^+) and the aromatic carbon atoms. This approach resembles the way we introduced the additional set of B_{ij} parameters. To fit the A_{ij} of charged N and aromatic carbon atoms we calculated single point energies for complexes of benzene with methylamine, tetramethylammonium cation, amidinium CH_3-$C^+(NH_2)_2$, and guanidinium. The cationic $N^+ sp^3$ species were placed in the middle of the benzene ring facing to it with the charged nitrogen. The N^+ sp^2 molecules were oriented coplanar to the benzene ring with the resonating carbon atom above the center. Then the molecules were shifted along the normal of the aromatic ring plane. The distances (in Å) between the aromatic ring plane and the charged nitrogen (the guanidine group plane) as well as the fitted energy corrections (kcal mol^{-1}) are listed in Table 2.

The accuracy of the fitting of cation-π interactions grows as the molecules move closer to each other. For the compact RNH_3^+ group the correction is almost sufficient, but it is very difficult to improve accuracy for the larger tetramethyl ammonium cation by the variation of only one atom repulsion exponent. The guanidinium and amidinium groups, having three and two nitrogen atoms respectively, fit perfectly at the distances smaller than 6 Å. As for longer distances, i.e., 5 Å and longer, the error is not so important, because the values calculated in vacuum have to be scaled for drug-design applications by the dielectric constant of water (i.e., 78 times reduced). As YFF1 model cation-π interactions as some kind of a van-der-Waals term, they are

Table 2 Cation-π interaction energies in kcal mol^{-1}

r, Å	$CH_3-NH_3^+$			$(CH_3)_4-N^+$			$CH_3-C(NH_2)_2^+$			$C(NH_2)_3^+$		
	QC	FF	FF$_0$	QC	FF	FF$_0$	QC	FF	FF$_0$	QC	FF	FF$_0$
2	70.53	1469	2731				310.72	37402	40327	193.58	2007	974
3	-13.04	-9.18	22.88	120.45	889.22	921.25	9.82	9.57	53.22	0.88	0.64	-11.67
4	-8.59	-0.95	0.69	-5.63	-0.04	1.60	-6.38	-6.14	-4.08	-5.25	-5.31	-5.83
5	-4.64	-0.14	0.01	-6.21	-1.08	-0.94	-3.97	-1.77	-1.58	-3.26	-1.97	-2.01
6	-2.73	-0.04	-0.02	-3.41	-0.40	-0.38	-2.38	-0.64	-0.61	-2.05	-0.77	-0.78
7	-1.74	-0.01	-0.01	-2.03	-0.16	-0.16	-1.54	-0.27	-0.27	-1.38	-0.34	-0.34
8	-1.17	-0.01	0.00	-1.32	-0.07	-0.07	-1.06	-0.13	-0.13	-0.98	-0.16	-0.16
9	-0.83	0.00	0.00	-0.91	-0.04	-0.04	-0.76	-0.07	-0.06	-0.91	-0.04	-0.04

QC is the quantum chemical dimerization energy calculated at MP2/6-31 G*, FF is the YFF1 energy corrected by the additional A_{ij} parameter for the aromatic carbon – charged nitrogen pair, and FF$_0$ is the YFF1 dimerization energy without correction

not scaled. Thus this cation-π interaction model suggests a polar (aqueous) environment, if the distance between the resonance system and the cation become 5.0 Å or more, as well as corrects interactions at closer distances reasonably well. The A_{ij} corrections for cation-π interactions in YFF1 (which covers only RNH_3^+ and $RC(NH_2)_2^+$) are given in Table S2 of Supporting information.

Validation with heterodimers

To evaluate the force field accuracy for configurations not present in the training set, we formed a validation dataset of

Fig. 3 Force field vs. quantum chemical dimerization energies (kcal mol^{-1})·of the heterodimeric validation set

244 heterodimeric configurations using monomers used in the parameterization. In order to maximize the variability of the validation set we used each monomer only once (488 unique monomers were randomly selected and coupled). Heterodimer energies were calculated at both the *ab initio* (MP2/6-31 G* with a BSSE correction) and the force field level (see Fig. 3). The average unsigned error for the validation set is 0.94 kcal mol^{-1} (the average square error is 1.71 kcal^2 mol^{-2}), which is even less than those of the training set. The average dimerization energy of the validation set is 7.43 kcal mol^{-1} (the average square is 58.34 kcal^2 mol^{-2}), which is even larger than those for the training set (see Table S3 for details). These values attest a satisfactory validation of the fitted parameters.

United-atom model

Because of the accuracy reasons, YFF1 is designed as an all-atom force field. However, we have also fitted a united-atom version, as it may be desirable to speed up some particular applications. Here, we however preserve aromatic hydrogen atom types, as they are required for a proper treatment of cation-π interactions in the YFF (otherwise cations would be able to be attracted to the tangential side of aromatic rings). Thus four new atom types were added: CH1 (aliphatic carbon with one hydrogen atom attached), CH2 (aliphatic carbon with two hydrogen atoms attached), CH3 (aliphatic carbon with three hydrogen atoms attached), CR (an sp^2 coordinated carbon with one or two hydrogen atom attached). To fit parameters of these atoms we neither performed the iterative FF convergence nor parameterized a new charge model. For the Coulomb interactions we simply sum up charges of all hydrogen atoms to condense them onto the corresponding carbon atom. For the Lennard-Jones parameters we used a closed all-atomic set of homodimeric configurations and
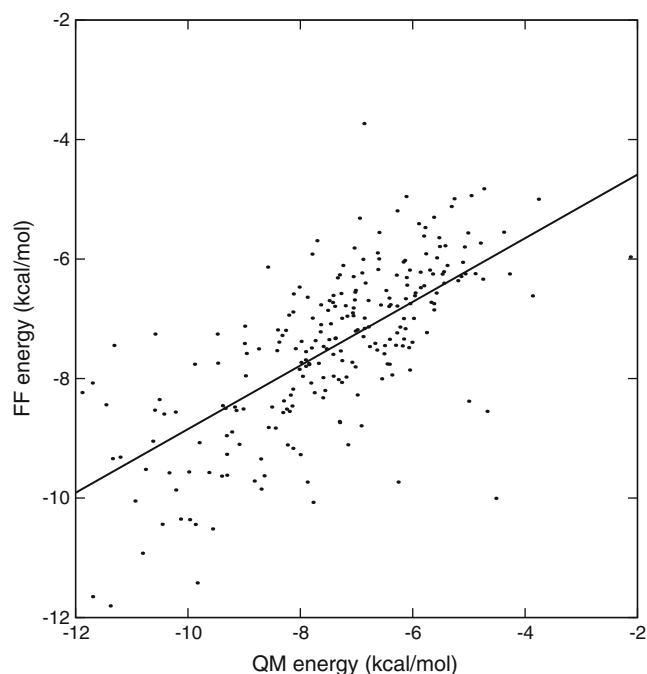
found parameters that were the best in interpolating the quantum chemical energies. Parameters of all atoms were taken from the optimized all-atom YFF1, thus we fitted only repulsive and attractive constants for four new atoms as well as the B_{ij} corrections matrix for all interactions between the united atom types. Simplex method was used for fitting, as described above. As expected, a small drop of accuracy comparing to the all-atom model was found. The average unsigned error after deleting 0.5 % of strong outliers for the united-atom model was 1.19 kcal mol^{-1} and the average square error as 3.25 kcal^2 mol^{-2}. The best fitted parameters for these united-atom types are listed in Table S1, whereas those for the correction B_{ij} matrix in Table S2 of the Supporting information.

Conclusions

In this work we report a nonbonded parameterization for the YFF1 force field. A training set of *ca.* 2,400 molecules (represented by *ca.* 6,600 configurations) was processed to calculate homodimerization energies at the MP2/6-31 G* level of theory with the Boys-Bernardi counterpoise BSSE correction. The new parameterization reproduces the high level quantum chemical energies with the average unsigned error as small as 1.10 kcal mol^{-1}. A validation set of 244 heterodimer configurations demonstrated an even lower unsigned average error of 0.94 kcal mol^{-1}. YFF1 is furnished with simple harmonic bonded terms, the 6–12 Lennard-Jones potential, and the classic Coulomb potential. YFF1 also has an improved nonbonded parameterization due to a special treatment of attractive constants of the most frequently occurred atomic pairs. Also special parameters are used for pair-wise repulsive exponents to model cation-π interactions. The proposed model of cation-π interactions works well at the distances of 2.0 - 5.0Å, but somewhat underestimates energies at larger distances. Parameterization of the Coulomb interactions was done previously by fitting against dipole and quadruple moments calculated at HF/6 G-311** theory level. Hard bonded parameters are derived from geometrical distortions calculated at the PM6 semiempirical method. Soft bonded parameters are entirely borrowed from MMFF94 where they were fitted to reproduce MP2/6 G-31* conformations energies. The intended primary functionality of YFF1 is as an all-atomic force field, but a united-atom model was parameterized as well.

Our parameterization suggests YFF1 as a good choice for fast but accurate evaluation of potential energies in configuration scanning routines of molecular docking software. It can also be used to minimize compounds geometry in gas phase. It is of interest for constructing of semi-explicit solvent shells (where the first water layer

molecules are placed explicitly, while all other are treated implicitly) around complexes being docked. Particularly, YFF1 is implemented as a part of our in-house Y_DOCK software for molecular docking, which will be reported separately.

Appendix A

FKCM (fast Kirchhoff charge model) formalism

Given a molecule of n atoms joined with m bonds. The internal structure operator matrix A is formed as: $A = I + \sum b_{ij} u_{ij} \otimes u_{ij}$, where I is the $n \times n$ identity matrix, b_{ij} is the bond weight (inverse bond hardness), $\otimes$ denotes tensor product, and u_{ij} is an n-vector such that $u_k = -\delta_{kj} + \delta_{ki}$ (i.e. its k-th element is -1 if $k=j$, $+1$ if $k=i$ and 0 otherwise). The KCM equation now reads:

$$A^{-1} \cdot \chi = q + \chi \tag{a.1}$$

where q is a charge vector and χ is an electronegativity vector. The solution of (a.1) is:

$$q = \left(A^{-1} - I\right) \cdot \chi, \tag{a.2}$$

If a molecule does not have any virtual counter-ions, then (a.2) is a straightforward solution. Otherwise operator A has to be perturbed in order to satisfy formal counter-ions charges $\pm Ze$, where Z is counter-ion's core charge. For once perturbed A corresponding b_{il} value (where l denotes the counter-ion index in charges vector) can be calculated analytically using the Sherman-Morrison formula:

$$(A + b_{il}u_{il} \otimes u_{il})^{-1} = A^{-1} - \frac{b_{il}A^{-1}.u_{il} \otimes u_{il}^T.A^{-1}}{1 + b_{il}u_{il}^T.A^{-1}.u_{il}} = A^{-1} - \frac{b_{il}v_{il} \otimes v_{il}}{1 + b_{il}u_{il}^T.v_{il}} \tag{a.3}$$

where $v_{il} = A^{-1}.u_{il}$. The final perturbation equation is obtained by setting Q_l as a formal charge of the virtual counter-ion and substituting (a.3) into (a.1):

$$A^{-1}\chi - \frac{b_{ii}v_{il} \otimes v_{il}.\chi}{1 + b_{il}u_{il}^T.v_{il}} = \chi_l - \frac{b_{il}K_{il}}{1 + b_{il}k_{il}} = Q_l + \chi_l, \tag{a.4}$$

$$\frac{b_{il}K_{il}}{1 + b_{il}k_{il}} = -Q_l.$$

because the lth row (column) of A and (A^{-1} as well) consists of only diagonal element 1, $k_{il} = u_{il}^T.v_{il} = 1 + v_{il}[i]$, (where

square brackets denote the ith component of v_{il} vector) and $K_{il} = v_{il}^T \otimes v_{il}.\chi = \chi_l - v_{il}^T.\chi$ again because of only unit diagonal element of lth row in A. The solution of (a.4) is then:

$$b_{il} = -\frac{Q_l}{K_{il} + k_{il}Q_l} = -\frac{Q_l}{\chi_l - v_{il}.\chi + Q_l(1 + v_{il}[i])}. \qquad (a.5)$$

For multiple virtual bonds in a molecule solution (a.5) should be applied to all virtual bonds in the system (*i.e.,* scales linearly) to calculate initial once perturbed guesses. In case of multiple bonds joined to the same virtual counter-ion it is recommended to use

$$Q` = \frac{Q}{m_{vi}},$$

where m_{vi} is the number of bonds attached to the given counter-ion (the approximation of an equivalent charge transfer). Initial guesses are then iteratively optimized to cope with the cross-perturbations in virtual hardnesses. Such optimization can be carried out with any suitable routine, particularly, l-BFGS method [42] is recommended to optimize initial perturbation guesses for cross-perturbation caused by multiple virtual hardness of the big protein molecule. Gradient of square charge's error over virtual hardnesses (derived from a.5) is only needed for l-BFGS routine:

$$A(b)^{-1}.\chi = Q` + \chi \xrightarrow{Q` \to Q} e(b) = \sum \left(A(b)^{-1}.\chi - \chi - Q \right)^2 \to 0,$$
$$\frac{\partial e_i}{\partial b_l} = -2((A^{-1}.\chi)_i - \chi_i - Q_i)A^{-1}\frac{\partial A}{\partial b_l}A^{-1}.\chi = -2\sqrt{e_i}A^{-1}(u_{il} \otimes u_{il})A^{-1}.\chi,$$
$$\frac{\partial e_i}{\partial b_l} = -2(q_l - \chi_l - Q_l) \cdot v_l[n+l] \cdot (v_l \bullet \chi), \qquad (a.6)$$

where $q_l = (A^{-1}.\chi)[n+l]$.

Appendix B

Fit of bonded interactions

PM6 semi-empirical method was a method against which parameters of hard bonded (HB) terms were fitted. Small geometrical distortions were used as a source of parameterization. As distortions are small enough, distances between atoms vary insignificantly and, hence, the non-bonded and soft bonded (SB) energies vary negligible enough to be treated as constant. Consequently, energy differences generated by small geometrical distortions can be considered as those caused by HB only. The deformation energy thus can be expressed as the second term of Taylor's series of geometrical distortions (bond lengths r, normal and improper angles, α and φ):

$$U_b \approx \frac{1}{2}\sum \frac{\partial^2 U}{\partial |r|^2}(r-r_0)^2 + \frac{1}{2}\sum \frac{\partial^2 U}{\partial \alpha^2}(\alpha - \alpha_0)^2 + \frac{1}{2}\sum \frac{\partial^2 U}{\partial \varphi^2}\varphi^2, \qquad (b.1)$$

where U is the potential energy, r is the bond vector, α are the bond angles and φ are the improper angles, the lower index 0 denotes the value in an energy minimum, and U_b is the distortion energy. Small geometrical distortions were generated by shifting each atomic coordinate away from the optimized geometry by a randomly chosen vector of the 0.05Å length. Relative energies of distorted geometries were then calculated using MOPAC2009 software. These relative distortion energies can be written as U_d-U_0 = U_b, where U_0 is energy of minimum and U_d is the energy of distorted geometry. As all energies in (b.1) depend linearly on the second energy derivatives over geometrical changes, the equation can be rewritten in the matrix form as:

$$A.x = y \qquad (b.2)$$

Table 3 Unsigned average deviation in HB from PM6 local minima

	Angles	H	C	N	O	F	P	S	Cl	Br	I
					Unsigned average deviations						
H			0.55	1.03	1.56		0.21	0.11			
C	1.25	0.55	1.25	1.51	0.98	0.34	1.15	1.49	0.58	0.97	0.87
N	1.97	1.03	1.51	1.26	0.46		1.48	1.42	0.02		
O	1.41	1.56	0.98	0.46	0.07		2.28	0.52			
F			0.34								
P	4.14	0.21	1.15	1.48	2.28			0.67			
S	1.63	0.11	1.49	1.42	0.52		0.67	0.73			
Cl			0.58	0.02							
Br			0.97								
I			0.87								

Deviation of angles (in degreases) and bonds (in pm) for different pairs of chemical atoms

Here A is a matrix of squares of geometric distortions (ith row describes geometry distortions of an ith compound of the training set), x is the vector of second energy derivatives over geometric distortions (i-th element of x means the ith FF parameter), and y is the vector of distortion energy calculated using PM6 (ith element of which means distortions of the ith compound).

System (b.2) was solved by the least squares method. During visual inspection of results we found that many of the rare parameters (encountered in less than 1 % of molecules) are often negative and/or unreasonably high in their absolute values. To cope with such artifacts parameters were manually adjusted: all outliers were set to corresponding values of carbon parameters (bonds to spring constant of C-C bond, angles to C-C-C angle and improper dihedrals to C-C-C-C). Then l-BFGS method was applied to optimize discrepancy between distortion energies:

$$s = (y - A\widetilde{x})^2, \quad \widetilde{x}_i = |x_i|. \tag{b.3}$$

l-BFGS routine had improved squared error value of (b.3) from the initial $1.43 \cdot 10^6$ to the final $5.69 \cdot 10^4$ kcal mol^{-1} per set in $2 \cdot 10^4$ steps. Final correlation between Ax product and energy distortions was as high as 0.97. HB parameters (stored in Tables S4, 5 and 6 of Supporting information) are: fitted average bond spring constant is ~240 kcal mol$^{-1} \cdot$Å^{-2}, average bond angle force constant is ~0.05 kcal mol$^{-1} \cdot$deg^{-2} and average improper angle parameter is ~0.0025 kcal mol$^{-1} \cdot$deg^{-2}. Average distortion energy over the set is +29.35 kcal mol^{-1}.

Parameterization of torsion barriers was perfectly made by Halgren and coworkers against conformation energies calculated at MP2/6-31 G* level of theory. Thus YFF1 entirely borrowed torsions parameterization and energy terms from MMFF94. For each of MMFF94 torsion angles the most appropriate atom types of YFF1 was substituted manually. Torsion energy expression of MMFF94 is taken into YFF1 without charges. The resulting YFF1's SB parameters are given in Tables S7 of the Supporting information.

To validate bonded interactions in YFF1 we re-minimized to nearest minima all monomers whose PM6 geometry has gradient norm less than 2 Hartree/Bohr at MP2/6 G-31* theory level (total 10438 compounds). Table 3 illustrates how far YFF1 optimized geometries in gas phase are comparing to those of PM6. Average unsigned error in bonds length is ~0.015 A, average unsigned error in angles is ~2 degrees, improper torsions are always very close to zero (data not shown), average root mean square deviation between structures at PM6 minima and YFF1 is only 0.04087 A^2. Although YFF1

geometry is somewhat robust it is sufficient for drug design purposes.

References

1. Allinger NL (1977) J Am Chem Soc 99:8127–8134
2. Allinger NL, Yuh YH, Lii JH (1989) J Am Chem Soc 111:8551–8565
3. Lii JH, Allinger NL (1989) J Am Chem Soc 111:8566–8575
4. Lii JH, Allinger NL (1989) J Am Chem Soc 111:8576–8582
5. Jorgensen WL, Tirado-Rives J (1988) J Am Chem Soc 110:1657–1666
6. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 118:11225–11236
7. Schuler LD, Daura X, Gunsteren WF (2001) J Comput Chem 22:1205–1218
8. Swart M, Duijnen PT (2006) Mol Simul 32:471–484
9. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) J Comput Chem 24:1999–201
10. Anisimov VM, Lamoureux G, Vorobyov IV, Huang N, Roux B, MacKerell AD (2005) J Chem Theory Comput 1:153–168
11. Burkert U, Allinger NJ (1982) Molecular Mechanics. Am Chem Soc, Washington
12. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) J Comput Chem 7:230–252
13. McCammon JA, Harvey SC (1987) Dynamics of proteins and nucleic acids. Cambridge University Press, Cambridge
14. Cornell WD, Cieplak P, Bayly IC, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179–5197
15. Baker CM, MacKerell AD (2010) J Mol Model 16:567–576
16. Kollman PA, Case DA (2003) In: Burger's Medical Chemistry and Drug Discovery, 6th edn, vol 1, Drug Discovery. John Wiley &Sons, New York
17. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E (2010) Guvench O, Lopes P, Vorobyov I, MacKerell AD. J Comput Chem 31:671–690
18. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) J Comput Chem 25:1157–1174
19. Maple JR, Hwang MJ, Stockfisch TP, Dinur U, Waldman M, Ewig CS, Hagler AT (1994) J Comput Chem 15:162–182
20. Halgren TA (1996) J Comput Chem 17:490–519
21. Charifson PS, Kunz ID (1997) In: Practical Application of Computer-Aided Drug Design. Dekker, New York
22. Yakovenko OYA, Oliferenko AA, Bdzhola VG, Palyulin VA, Zefirov NS (2008) J Comput Chem 29:1332–1343
23. Wang J, Wang W, Kollman PA, Case DA (2006) J Mol Graph Model 25:247–260
24. Aqvist J, Medina C, Samuelsson JE (1994) Prot Eng 7:385–391
25. Stjernschantz E, Marelius J, Medina C, Jacobsson M, Vermeulen NPE, Oostenbrink C (2006) J Chem Inf Model 46:1972–1983
26. Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001) J Comput Aided Mol Des 15:411–428
27. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Comput Chem 19:1639–1662
28. Rarey M, Kramer B, Lengauer T, Klebe G (1996) J Mol Biol 261:470–489
29. Jain AN (2003) J Med Chem 46:499–511
30. Jones G, Willett P, Glen RC, Leach AR, Taylor RJ (1997) Mol Biol 267:727–748
31. Abagyan RA, Totrov MM, Kuznetsov DA (1994) J Comput Chem 15:488–506

32. http://www.kressworks.com/kressworksorg/Quantum_Chemistry/Potential_Energy_Surfaces/water_dimer/water_dimer_results.html
33. Kaminski GA, Stern HA, Berne BJ, Friesner RA (2004) J Phys Chem 112:621–627
34. Giese TJ, York DM (2004) J Chem Phys 120:9903–9906
35. Sherman J, Morrison WJ (1950) Ann Math Stat 21:124–127
36. Granovsky AA (2010) Firefly version 7.1.G, http://classic.chem.msu.su/gran/firefly/index.html
37. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus TL, Dupuis M, Montgomery JA (1993) J Comput Chem 14:1347–1363
38. Boys SF, Bernardi F (1970) Mol Phys 19:553–566
39. http://zinc7.docking.org/
40. Stewart JJP (2007) J Mol Model 13:1173–1213
41. Nelder JA, Mead R (1965) Comput J 7:308–313
42. Liu DC, Nocedal J (1989) Math Program 45:503–528

ORIGINAL PAPER

Developing consensus 3D-QSAR and pharmacophore models for several beta-secretase, farnesyl transferase and histone deacetylase inhibitors

Hsin-Yuan Wei · Guan-Ju Chen · Chih-Lun Chen · Thy-Hou Lin

Abstract Three consensus 3D-QSAR (c-3D-QSAR) models were built for 38, 34, and 78 inhibitors of β-secretase, histone deacetylase, and farnesyltransferase, respectively. To build an individual 3D-QSAR model, the structures of an inhibitor series are aligned through docking of a protein receptor into the active site using the program GOLD. CoMFA, CoMSIA, and Catalyst are then performed for the training set of each structurally aligned inhibitor series to obtain a 3D-QSAR model. Since the consensus in features identified is high for the same pharmacophore features selected for building a 3D-QSAR model by a 3D-QSAR method, a c-3D-QSAR model for each inhibitor series is constructed by combining the pharmacophore features selected for building the 3D-QSAR model using the SYBYL spread sheet and PLS module. Each c-3D-QSAR pharmacophore model built was examined visually and compared with that obtained by simultaneous mapping of the corresponding 3D-QSAR pharmacophores built onto a selected inhibitor structure. It was found that the c-3D-QSAR model built for an inhibitor series improves not only the overall prediction statistics for both training and test sets but also the prediction accuracy for some less active inhibitors of the series.

H.-Y. Wei · G.-J. Chen · C.-L. Chen · T.-H. Lin (✉)
Institute of Molecular Medicine and Department of Life Science,
National Tsing Hua University,
Hsinchu, 30013, Taiwan, Republic of China
e-mail: thlin@life.nthu.edu.tw

H.-Y. Wei
e-mail: biotiger64@hotmail.com

Introduction

Alzheimer's disease (AD) is a debilitating disease that patients often suffer for several years or even longer; hence, treatment is costly [1]. It is well established that AD stems from the accumulation of 40/42-residue amyloid β-peptide (Aβ), leading to the formation of insoluble plaques in the brain [2]. Aβ is produced through first cleavage of a membrane amyloid precursor protein (APP) by a protease known as β-secretase, which has been identified and designated as the membrane-anchored aspartic protease BACE-1 [2–4]. The matured Aβ is then generated by a second cleavage at the C-terminus by γ-secretase [2, 3]. It has been shown that BACE-1 knockout mice are unable to produce Aβ either from endogenous APP or a mutant human APP transgene [5]. This establishes BACE-1 as a potential drug target for developing therapeutic agents for treating this dreadful and prevalent disease. Recently, Iserloh et al. [6] have performed extensive structure–activity relationship studies on both pyrrolidine and piperidine classes of compounds as BACE-1 inhibitors. Lindsley et al. [7] designed and synthesized some tertiary macrocyclic carbinamines as BACE-1 inhibitors. Moreover, a series of isophthalamide derivatives have been synthesized and evaluated by Stachel et al. [8], and were shown to be potent and selective BACE-1 inhibitors. Most BACE-1 inhibitors reported to date are of the peptidomimetic type and have been reviewed in the literature [9]. However, Geschwindner et al. [10] have used a fragment-based technique to find some novel small-molecule BACE-1 inhibitors. By applying

a tethering technique to the X-ray crystal structures of some BACE-1 mutant-disulfide conjugates, Yang et al. [11] synthesized a series of nonpeptidic BACE-1 inhibitors based on the central aminobenzylpiperidine moiety. Two well-known three-dimensional quantitative structure-activity relationship (3D-QSAR) methods, namely comparative molecular field analysis (CoMFA) [12] and comparative molecular similarity indices analysis (CoMSIA) [13] were used by Pandey et al. [14] with 43 hydroxyethylamine derivatives as potent BACE-1 inhibitors to build some pharmacophore models for the enzyme.

Zinc (Zn^{2+})-dependent histone deacetylases (HDACs) are found to be overexpressed in several cancer types [15]. The acetylation of lysine residues in nucleosomal histones is mediated by HDACs [15–17]. Therefore, these enzymes are responsible for controlling gene expression and cell cycle progression [16, 17]. Inhibition of HDACs by some antitumor agents such as trichostatin A (TSA) [18], suberoylanilide hydroxamic acid (SAHA) [19], PXD-101 [20], and MS-275 [21] is found to inhibit cell growth, induce terminal differentiation in tumor cells, and prevent the formation of malignant tumors. There is also evidence indicating that HDAC inhibitors can be used as effective therapeutic agents to treat some inflammative and neurodegenerative diseases [22]. HDACs are classified structurally and functionally into several classes, namely class I (HDAC 1–3 and 8), class II (HDAC 4–7, 9 and 10), class IV (HDAC 11), and class III [23]. The class III HDACs are also zinc-dependent but are structurally distinct from the other classes and require the cofactor NAD^+ for their deacetylase function. Most of the HDAC inhibitors published to date, such as panobinostat, belinostat and vorinostat (formerly known as SAHA [24], and approved by FDA for the treatment of cutaneous T-cell lymphoma), contain a hydroxamic acid group. There are also non-hydroxamic acid-based inhibitors being developed that include 2-aminophenylamide MS-275 [25] and dithiol FK228 [26]; both are currently in clinical trials. Most of these inhibitors hit a subset of both class I and II or, as in the case of 2-aminophenylamides [25] show selectivity versus class I HDACs. However, all of these inhibitors elicit similar adverse effects, mainly fatigue, nausea, vomiting, and diarrhea, that become dose-limiting in clinical trials. A QSAR study published by Wang et al. [27] on some TSA- and SAHA-like hydroxamic-acid-based compounds has shown that the shape and area of these molecules are important for their biological activity. Using a data set of 124 compounds, Xie et al. [28] have shown that van der Waals surface area and hydrophobicity are important parameters required for biological activity. To build a 3D pharmacophore model, Juvale et al. [29] have employed both CoMFA and CoMSIA methods on a set of 40 hydroxamic acid analogues and found that the electrostatic interactions

between the hydroxamic acid group and enzyme residues play a major role in binding to the HDAC active site.

As a GTP-binding protein, Ras plays a key role in the cell signal transduction pathways that mediate cell growth and lead to cellular transformation and uncontrolled proliferation [30]. Mutant Ras genes are often found in human tumors [31], suggesting that inhibition of Ras function might provide an effective anticancer therapy. Ras proteins are produced as cytoplasmatic precursor proteins and several posttranslational modifications are required to acquire their full biologic function [32, 33]. The most important step among these is the transfer of a farnesyl group from farnesyldiphosphate (FPP) to the cysteine residue of a CAAX (C represents cysteine, A represents an aliphatic amino acid, and X represents any amino acid) motif on the protein [34]; this reaction is catalyzed by zinc metalloenzyme farnesyltransferase (FTase). FTase is a zinc heterodimeric metalloenzyme consisting of two subunits of 48 kDa α [35] and 46 kDa β [36]. The zinc ion lies in a hydrophilic surface formed by the α-subunit and a deep cleft of the β-subunit surrounded by some aromatic residues [37]. The FTase substrates have been treated as models for designing some selective FTase inhibitors. Most of the FTase inhibitors described in the literature are peptidomimetics resembling the CAAX-tetrapeptide recognition sequence of farnesylated proteins [38, 39]. However, the design of FTase inhibitors has evolved from early thiol-containing peptidimimetics to recent non-thiol compounds [38–42]. The most frequently used replacements for cysteine are the nitrogen-containing heterocycles, where the nitrogen is presumably interacting with the zinc ion in a similar way as to the cysteine thiol group [41, 42]. However, a series of potent FTase inhibitors have been synthesized through combining the nitrogen-containing heterocycles with various carboxyl terminus mimics such as the terminal phenylsulfonyl group [43]. The 4-cyanobenzyl group on the imidazole ring is crucial for binding activity and is found to reach into a high-affinity aromatic binding pocket that is otherwise inaccessible by the tetrapeptides. Moreover, substitution of the 4-cyanophenyl ring with an additional 3-aryloxy group with or without concomitant removal of the cyano group also creates a series of potent FTIs [44, 45]. Recently, Puntambekar et al. [37] derived some 3D-QSAR models using both CoMFA and CoMSIA methods for 38 3-aminopyrrolidinone derivatives, 46 2-amino-nicotinonitriles, and 35 1-aryl-1′-imidazolyl methyl ethers. Xie et al. [46] have applied the same 3D-QSAR techniques plus the molecular docking program GOLD to a highly diverse set of 192 Abbott-initiated imidazole-containing FTase inhibitors. Moreover, the Catalyst HypoGen program has been used by Equbal et al. [47] to build some 3D pharmacophore models from 22 FTase inhibitors.

In this report, we have constructed some individual 3D-QSAR models by applying the CoMFA, CoMSIA and Catalyst methods to three series of inhibitors targeted to BACE-1, HDAC, and FTase, respectively. Both CoMFA [12] and CoMSIA [13] methods are now widely used as standards for building 3D-QSAR models. The most important perspective given by the CoMFA model built using partial least square (PLS) analysis [48] are contour maps showing favorable and unfavorable regions for electropositive or electronegative, or favorable and unfavorable regions for steric substituents in certain positions. Prediction for the test set can be made either by a qualitative inspection of these contour maps or, in a quantitative manner by calculating the fields of these molecules and by inserting the grid values into the PLS model. The same grid constructed for CoMFA is also used for CoMSIA calculation. Besides the electrostatic (E) and steric (S) fields, CoMSIA also offers hydrophobic (H), hydrogen bond (H-bond) donor (D) and H-bond acceptor (A) field information. CoMSIA is thought to be less affected by changes in molecular alignment, and smoother contour maps are provided by using the Gaussian type distance dependence on the molecular similarity indices computed. Catalyst [49] is also a powerful pharmacophore building method employing up to 11 pharmacophore features, namely H-bond donor (HBD), H-bond acceptor (HBA), HBA lipid, hydrophobic (HY), hydrophobic aliphatic (HL), hydrophobic aromatic (HR), negative charge, negative ionizable, positive charge, positive ionizable, and ring aromatic (RA) in the computation. The activity data of the training set compounds are randomized and then validated by the catScramble module using the Fischer's randomization test [50] for generating some phony hypotheses and statistical costs using the same parameters used for generating the original ones. The original hypotheses are considered to be statistical significant only if their corresponding statistical costs and computed correlation are better than those computed for the randomized sets. We have devised a consensus scheme from all the three 3D-QSAR models built for each series of inhibitors. The rationale for designing a consensus 3D-QSAR (c-3D-QSAR) model is that most of the pharmacophore features generated by the 3D-QSAR methods will overlap if the ligand structures are generated correctly and aligned. On the other hand, different pharmacophore features obtained by each 3D-QSAR can be combined to supplement each other to give a c-3D-QSAR model from which more versatile pharmacophore features and better prediction accuracy may be derived. Each c-3D-QSAR model was constructed using the SYBYL PLS module and the corresponding c-3D-QSAR pharmacophore was generated automatically by the SYBYL program. Each of these c-3D-QSAR pharmacophores was

examined visually and compared with those obtained by mapping each 3D-QSAR pharmacophore onto a selected inhibitor structure. The procedures for constructing each 3D-QSAR and c-3D-QSAR model for each inhibitor series are detailed in the Materials and methods while the performance of each 3D-QSAR and c-3D-QSAR model built are compared and discussed in the Results and Discussion. The significance of building a c-3D-QSAR model is addressed in the Conclusion section.

Materials and methods

Data set and ligand preparation

The two-dimensional (2D) structures, biological activities (expressed in pIC_{50}), and original inhibitor identification numbers of 38 BACE-1 inhibitors synthesized by Freskos et al. [51, 52] are listed in Table S1.1 in the electronic supplementary material. Since there were repetitions in original identification numbers, we assigned each inhibitor studied a new inhibitor number #, and these new #s as listed in Table S1.1 are used throughout the report. The X-ray crystallography determined structure 2HM1, where **a5** was engulfed [52], was used as a structural template to build the structures for other inhibitors of the series. The structure of each of the other inhibitors of the series was constructed within the active site of 2HM1 by replacing the side chains of **a5** with other functional groups as described previously (Table S1.1) [53]. Hydrogen atoms were added for each structure. The pIC_{50} measured for these BACE-1 inhibitors ranged from 8.70 to >4.70 (Table S1.1). Each structure constructed was subjected to some energy minimization steps together with the receptor until a convergence of 0.05 kcal mol^{-1} Å was reached using the AMBER 7 F99 charges [54] implemented in the SYBYL 8.0 program [55]. Then, we docked ligand **a5** into the active site of BACE-1 using the GOLD V4.0 program [56] and computed the root-mean-square-deviation (RMSD) between the docked conformation and the X-ray structure. The major docking parameters, namely number of operation and population size, were set as 100,000 and 100, respectively. Each energy-minimized structure was subsequently docked into the same active site of BACE-1 using the same GOLD docking parameters. The AM1-BCC charges [57] computed by the AMBER 9.0 program [58] were deployed for each inhibitor. The 2D structures, pIC_{50}, original inhibitor identification numbers, plus our assigned #s for the 34 HDAC inhibitors synthesized by Scarpelli et al. [59] and Muraglia et al. [60] are listed in Table S1.2. The pIC_{50} measured [59, 60] for these inhibitors ranged from 8.15 to >5 (Table S1.2). The same structure preparation steps used for BACE-1

inhibitors were employed for the HDAC class II 4WT inhibitors except that the ligand structure TFG (designated as **b30-1** in Table S1.2) of 2VQJ [61] was used only as a template to construct all the inhibitor structures of the series and was not included in the 3D-QSAR analyses.

The total number of FTase inhibitors studied was 78, and the corresponding 2D structures, pIC_{50}, original inhibitor identification numbers, plus our assigned #s are listed in Table S1.3. The inhibitors synthesized by Tong et al. [44] were designated as the **c-#** series, while those synthesized by Wang et al. [45] were designated as the **d-#** series (Table S1.3). The pIC_{50} measured for these FTIs ranged from 9.80 to 7.08 [44, 45] (Table S1.3). The crystal ligand **c-19** (Table S1.3) of X-ray structure 1NI1 [44] was used as a structural template to build all the inhibitor structures of this series. The structure preparation steps used for this series were similar to those described for the BACE-1 inhibitors.

3D-QSAR models

The enantiomer type taken for each inhibitor is specified in Tables S1.1, S1.2 and S1.3. After all the structures were built, each series of inhibitors was divided into a training set and a test set while keeping the range of pIC_{50} measured in each subset roughly the same. The numbers of training set inhibitors allocated for the BACE-1, HDAC, and FTase series were 24, 19, and 43, respectively, while those allocated for the corresponding test sets were 14, 15, and 35, respectively. Before proceeding with the CoMFA and CoMSIA computations, all structures of training set inhibitors were aligned based on the docked conformations generated by the GOLD V4.0 program. The aligned structures of all the training set inhibitors of the BACE-1, HDAC, and FTase series are presented in Figs. S2.1, S2.2, and S2.3, respectively. The steric and electrostatic potential fields of CoMFA were calculated with the SYBYL 8.0 program using a regularly spaced grid of 2.0 Å. The grid was extended to 4 Å units beyond the van der Waals volume of each molecule in the X, Y, and Z directions. An sp3 carbon atom of radius of 1.52 Å and charge of +1.0 was used as a probe to calculate both steric and electrostatic interaction energies. The truncation for both steric and electrostatic contributions was set at ±30 kcal mol^{-1}. The electrostatic contribution at the grid intersections where maximum steric interactions were computed was ignored. Both CoMFA steric and electrostatic fields computed were scaled by the standard option given by the program. An sp3 atom of radius of 1.0 Å and charge of +1 was chosen as the probe for computing the CoMSIA similarity indices defined by Klebe et al. [13]. The attenuation factor α was set at 0.3. The CoMFA and CoMSIA descriptors were treated as independent variables while the measured pIC_{50}

was treated as the dependent variable in all the PLS regression analyses for deriving each 3D-QSAR model. The optimum number of components used to derive a nonvalidated model was defined as the number of components used to obtain the highest crossvalidated r^2 (q^2) and lowest standard error of prediction (SEP) computed. The goodness-of-fit of nonvalidated models was judged by the conventional computed correlation coefficient r^2, leave-one-out validated r^2 (q_{loo}^2), q^2, SEP, and F values. The nonvalidated analysis results were used to predict pIC_{50} values.

Generation of the pharmacophore hypotheses

The same structurally aligned training sets used to construct the CoMFA and CoMSIA 3D-QSAR models were also used to construct the pharmacophore hypotheses using the Catalyst 4.11 program [49]. All parameters used were default settings except for Unc (uncertainty); this parameter was set as 1.9 for more active inhibitors, namely $IC_{50} < 1$ nM, or 2.0 for less active ones where measured $IC_{50} \geq 1$ nM. The pharmacophore features selected for building the top hypotheses were HBA, HR, and HL for BACE-1; HBA, HR, and RA for HDAC; and HY, HR, and RA for FTase inhibitor series, respectively. These pharmacophore features were determined from a series of stepwise CoMSIA runs, namely a single field index, a combination of any two field indexes, a combination of any three field indexes, a combination of any four field indexes, and a combination of all field indexes were used. The best set of pharmacophore features was chosen as the set with the best q_{loo}^2, q^2, SEP, and F values obtained. The top ten scored hypotheses were then generated by the HypoGen module of the Catalyst 4.11 program [49] using the best pharmacophore features determined from the stepwise CoMSIA results for each inhibitor of each series. The goodness-of-fit of mapping a ligand structure onto a Catalyst pharmacophore feature of the top hypothesis, namely fit_{PF}, was computed by the Catalyst Citest module by setting the parameter Max omitted features as 1 (the default setting).

Generation of c-3D-QSAR

Each consensus 3D-QSAR (c-3D-QSAR) model was derived from the best CoMFA, CoMSIA, and Catalyst 3D-QSAR models built from the training set of each inhibitor series using the SYBYL PLS module. The same spreadsheet used for obtaining the best 3D-QSAR CoMFA and CoMSIA models for the training set inhibitors was opened for input for each Catalyst fit_{PF} value. For example, for the training set HDAC inhibitors, the four Catalyst fit_{PF} columns obtained, namely fit_{HBA1}, fit_{HBA2}, fit_{HR3}, and fit_{RA4}, were typed manually into the spreadsheet as the

last four columns. The consensus spreadsheet (c-spreadsheet) then consisted of nine columns, namely pIC_{50}, CoMFA, CoMSIA_S, CoMSIA_E, CoMSIA_H, Catalyst fit$_{HBA1}$, fit$_{HBA2}$, fit$_{HR3}$, and fit$_{RA4}$ (Fig. S1). The c-spreadsheet was analyzed by the SYBYL PLS module with options of crossvalidation or no validation to obtain the desired statistics. The pIC_{50} values for each test set were predicted via a no validation run on the corresponding c-spreadsheet. The major PLS analysis results were either outputted as a .lis file for reporting the statistics or a .pls file for predicting the pIC_{50} for the test set and viewing the consensus pharmacophore (c-pharmacophore) of the c-3D-QSAR model built.

Results and discussion

Determination of individual 3D-QSAR models

3D-QSAR models of BACE-1 inhibitors

A RMSD of 0.44 Å was obtained for docking inhibitor **a5** into the active site of 2HM1, indicating that the parameters

chosen for subsequent docking of each theoretically generated structure of the BACE-1 series into the same active site were adequate. The best GOLD-docked conformation of each inhibitor of the series was used for constructing all the individual 3D-QSAR models. As shown in Table 1 for the BACE-1 series, the following statistics, namely $q_{loo}^2 = 0.85$, $q^2 = 0.85$, $r^2 = 1.00$, SEP = 0.04, and F = 3,044, were obtained for the best CoMFA 3D-QSAR model. As judged by SEP, q^2, and F computed, no single field CoMSIA obtained was better than that of a combination of the S+H+E+A field indexes (Table S2.1). Next, no apparent improvement in CoMSIA statistics was obtained when all the five field indexes were used in the computation (Table S2.1). Therefore, the interaction of the BACE-1 inhibitors with their active site is best described by a combination of S+H + E+A field indexes (Table S2.1). The corresponding statistics obtained for this best CoMSIA 3D-QSAR model were $q_{loo}^2 = 0.86$, $q^2 = 0.86$, $r^2 = 0.98$, SEP = 0.18, and F = 356 (Table 1). Based on the best CoMSIA 3D-QSAR model built, the following pharmacophore features, namely HBA, HL, and HR were selected for constructing the Catalyst 3D-QSAR models. The HypoGen module of the Catalyst 4.11

Table 1 A summary of statistics obtained for each CoMFA, CoMSIA, and Catalyst 3D-QSAR and c-3D-QSAR model built for each training and test set inhibitor of each series

BACE-1 inhibitors		CoMFA	CoMSIA (S+E+H + A)	Catalyst	c-3D-QSAR
Training set	Leave one out q_{loo}^2	0.85	0.86	0.97	0.93
	Cross-validation q^2	0.85	0.86	0.97	0.92
	Conventional r^2	1.00	0.98	0.98	0.99
	Standard error	0.04	0.18	0.19	0.14
	Principal components	6	3	1	5
	F-value	3044	356	939	337
Test set	r^2	0.86	0.80	0.99	0.97
HDAC inhibitors		CoMFA	CoMSIA (S+E + H)	Catalyst	c-3D-QSAR
Training set	Leave one out q_{loo}^2	0.76	0.76	0.87	0.84
	Cross-validation q^2	0.75	0.76	0.87	0.81
	Conventional r^2	1.00	1.00	0.9	1.00
	Standard error	0.04	0.06	0.26	0.06
	Principal components	6	6	1	6
	F-value	1186	529	152	523
Test set	r^2	0.80	0.81	0.79	0.89
Test set	r^2	0.69	0.75	0.71	0.97
FTase inhibitors		CoMFA	CoMSIA (S+H+E)	Catalyst	c-3D-QSAR
Training set	Leave one out q_{loo}^2	0.66	0.69	0.80	0.75
	Cross-validation q^2	0.66	0.67	0.80	0.75
	Conventional r^2	0.98	0.98	0.90	0.97
	Standard error	0.09	0.10	0.26	0.11
	Principal components	5	5	1	6
	F-value	375	285	180	224
Test set	r^2	0.69	0.75	0.71	0.97

program was used to generate the desired pharmacophores automatically. The top ten scored hypotheses generated were exported and their validity examined using the catScramble module of the Catalyst program (Table S3.1). With a cost difference between null and total 1 being 74.3 and that between null and fixed 1 being 88.8, the top hypothesis generated meets the criteria of being a good hypothesis (Table S3.1). The corresponding statistics obtained for the Catalyst 3D-QSAR model built are $q_{loo}^2 = 0.97$, $q^2 = 0.97$, $r^2 = 0.98$, SEP = 0.19, and F = 939 (Table 1). Apparently, the statistics given by the Catalyst 3D-QSAR model are slightly better than those given by either CoMFA or CoMSIA 3D-QSAR (Table 1). The measured and predicted pIC_{50} values given by the CoMFA, CoMSIA and Catalyst 3D-QSAR models built for training set BACE-1 inhibitors are compared in Table 2 and Fig. S3.1. Apparently, both the CoMSIA and Catalyst 3D-QSAR models give some discrepancies in the predicted versus measured pIC_{50} values for inhibitor **a3**, **a21**, **a24**, **a26**, **a28** and **a29** and that given by all the three 3D-QSAR models built for inhibitor **a30**, **a31**, **a33**, and **a37** are also significant as well (Table 2).

The steric and electrostatic contours of the structure of **a5** depicted by CoMFA and CoMSIA 3D-QSAR models are compared in Fig. 1a–d. Apparently, the S-disfavored regions represented by yellow contours given by both CoMFA and CoMSIA 3D-QSAR models are different but some S features near P3 pocket represented by green contours are detected by both models (Fig. 1a,b). However, better agreement in the E features detected by both CoMFA and CoMSIA 3D-QSAR models can be seen in Fig. 1c,d, where favored regions for positive or negative charges, represented by blue or red contours, respectively, are detected for p2 and p2' or near p2 pockets [14]. The H features detected by the CoMSIA and Catalyst 3D-QSAR models are compared in Fig. 1e and f, respectively. The two H features represented by orange contours in the CoMSIA 3D-QSAR model near the p2' and p3 pockets are also detected by the Catalyst HR3 and HL features represented by cyan and blue spheres, respectively (Fig. 1e,f). The HBA feature represented by green spheres detected by the Catalyst 3D-QSAR is also in accord with the A feature represented by magenta contours detected by the CoMSIA 3D-QSAR model (Fig. 1e,f). Note that there are two white

Table 2 Actual (Act) and predicted (Pred) pIC_{50} values for the BACE-1 inhibitor training set derived from the best CoMFA, CoMSIA, and Catalyst 3D-QSAR models compared with those of the c-3D-QSAR model

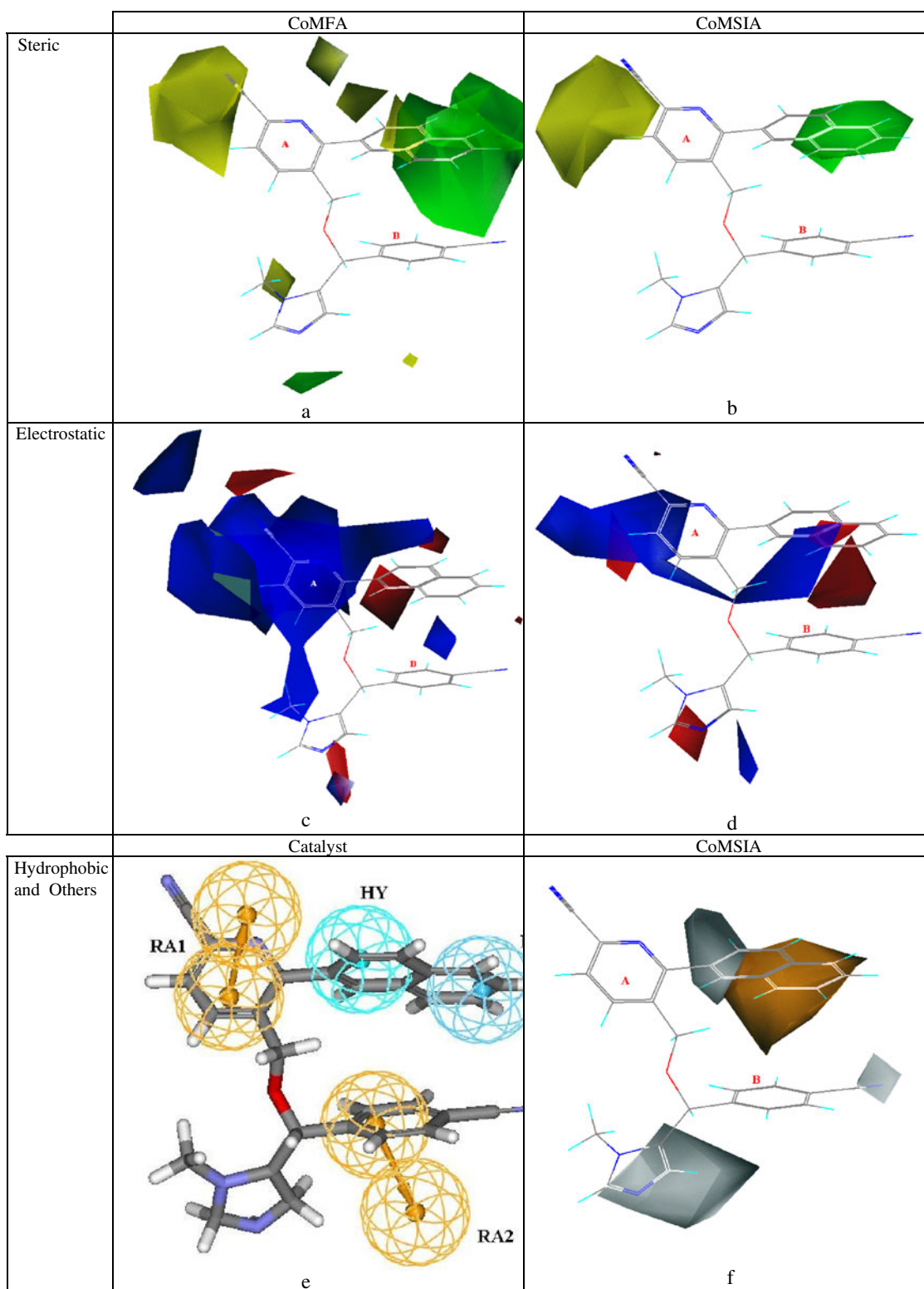| Inhibitor#	Act pIC_{50}	CoMFA Pred pIC_{50}	CoMSIA (S+E+H+A) Pred pIC_{50}	Catalyst (hypothesis Hypo1) Pred pIC_{50}	c-3D-QSAR Pred pIC_{50}
a3	8.70	8.71	8.53	8.31	8.59
a4	8.70	8.72	8.69	8.68	8.84
a5	8.70	8.69	8.70	8.48	8.48
a6	8.52	8.48	8.17	8.54	8.56
a7	8.52	8.50	8.47	8.39	8.44
a10	8.40	8.41	8.52	8.41	8.46
a11	8.40	8.41	8.21	8.41	8.39
a12	8.40	8.42	8.55	8.28	8.29
a13	8.40	8.39	8.40	8.35	8.35
a17	8.30	8.30	8.26	8.24	8.10
a18	8.30	8.32	8.32	8.36	8.21
a20	8.00	8.01	7.93	8.11	8.03
a21	8.00	7.97	8.30	8.28	8.23
a22	7.96	7.98	7.98	8.19	7.98
a24	7.82	7.82	7.91	8.08	8.03
a26	7.66	7.70	7.84	7.89	7.87
a28	7.28	7.25	7.39	7.17	7.33
a29	6.95	6.96	6.77	6.66	6.88
a30	6.32	6.25	6.42	6.59	6.29
a31	6.22	6.16	5.91	6.48	6.15
a33	5.44	5.41	5.37	5.62	5.48
a34	5.42	5.44	5.47	5.35	5.30
a36	5.28	5.27	5.28	5.25	5.26
a37	5.25	5.36	5.54	5.36	5.41
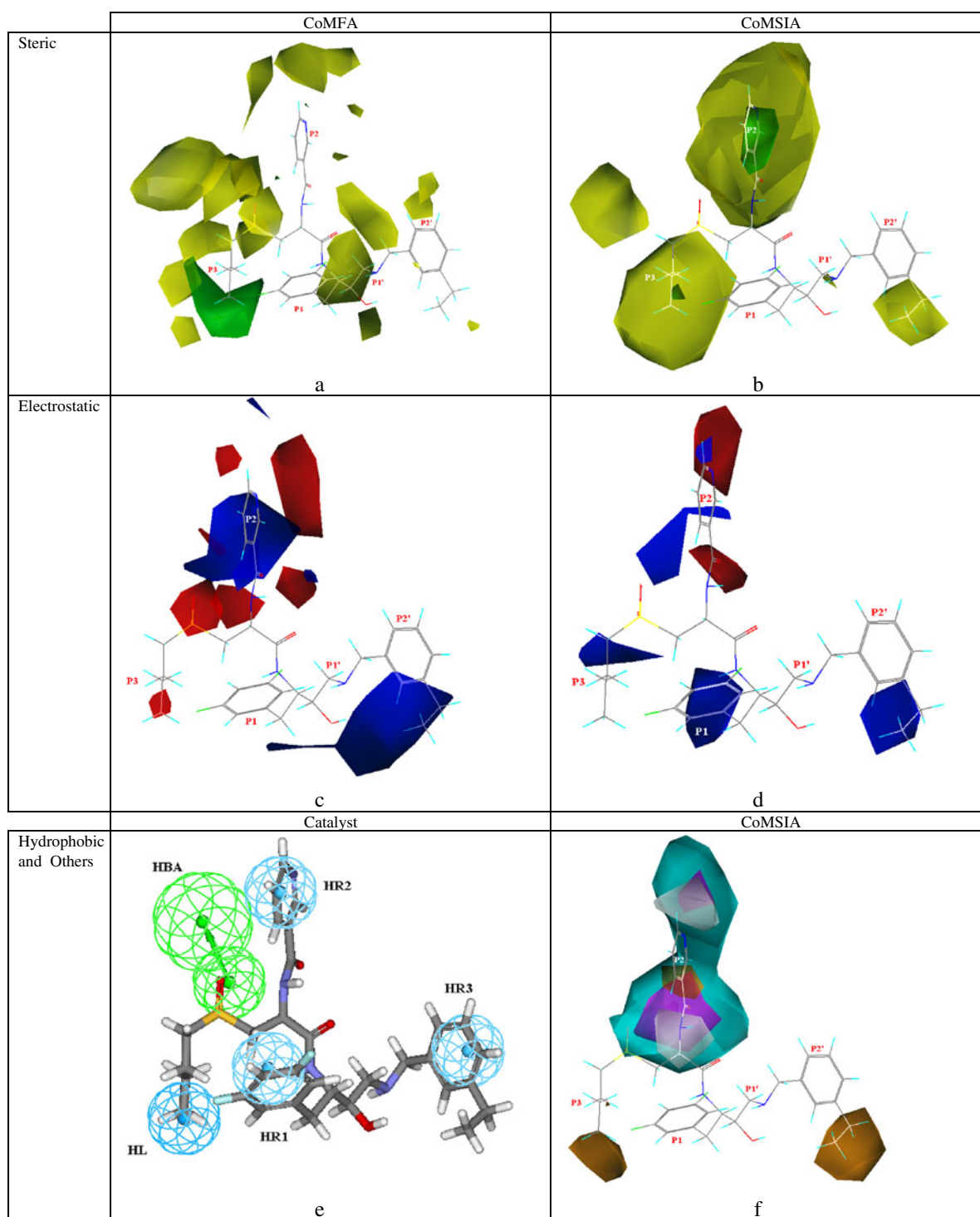
Fig. 1 a–f Pharmacophore features detected by the best CoMFA, CoMSIA, and Catalyst 3D-QSAR models built from the BACE-1 inhibitors training set. The structure of **a5** is mapped onto the steric pharmacophore features of each model. **a** CoMFA steric contours (*green* sterically favored, *yellow* sterically disfavored) of the best CoMFA 3D-QSAR model. **b** CoMSIA steric contours (*green* sterically favored, *yellow* sterically disfavored) of the best CoMSIA 3D-QSAR model. **c** CoMFA electrostatic contours (*blue* positive charge favored, *red* negative charge favored) of the best CoMFA 3D-QSAR model built. **d** CoMSIA electrostatic contours (*blue* positive charge favored, *red* negative charge favored) of the best CoMSIA 3D-QSAR model. **e** The structure of **a5** mapped onto the Catalyst 3D-QSAR pharmacophore features built from the BACE-1 inhibitors training set. Pharmacophore features: *cyan spheres* hydrophobic aromatic (HR), *blue spheres* hydrophobic aliphatic (HL), *green spheres* H-bond acceptor (HBA) features. **f** CoMSIA H and A contours (*orange* hydrophobic favored, *white* hydrophobic disfavored, *magenta* HBA favored, *cyan* HBA disfavored regions) of the best CoMSIA 3D-QSAR model. The structure of **a5** is mapped onto the CoMSIA H and A pharmacophore features

contours representing disfavored regions for H feature detected by the CoMSIA 3D-QSAR model near the p2 pocket, which are undetected by any Catalyst hydrophobic features employed (Fig. 1e,f). These comparisons show that the more versatile hydrophobic features provided by the Catalyst 3D-QSAR model may be used to supplement the H feature provided by the CoMSIA 3D-QSAR model.

3D-QSAR models of HDAC inhibitors

The RMSD computed for the GOLD-docked conformation of the 2VQJ ligand [61] b30-1, which is the X-ray structure of wild type HDAC 4 protein, is 0.5 Å. The best CoMFA 3D-QSAR model built from the training set of the series gave the following statistics: $q_{loo}^2 = 0.76$, $q^2 = 0.75$, $r^2 = 1.00$, SEP = 0.04, and F = 1,186 (Table 1). The best CoMSIA 3D-QSAR model was also constructed in a stepwise manner by a combination of S+H+E filed indexes (Table S2.2) and the corresponding statistics obtained were: $q_{loo}^2 = 0.76$, $q^2 = 0.76$, $r^2 = 1.00$, SEP = 0.06, and F = 529 (Table 1). The best CoMSIA 3D-QSAR model obtained gives us clues to select the following pharmacophore features, namely RA, HR, HBA and HBA, to build the Catalyst 3D-QSAR model using the Catalyst HypoGen module. As validated by the catScramble module, the cost difference obtained between the null and total one is 66.6,

while that between null and fixed one is 76.5 for the top hypothesis generated (Table S3.2). The corresponding statistics in predicting pIC_{50} for the HDAC inhibitors training set given by the top hypothesis are listed in Table 1. Again, both q_{loo}^2 (0.87) and q^2 (0.87) obtained by the Catalyst 3D-QSAR model are slightly better than those by either CoMFA or CoMSIA 3D-QSAR (Table 1). However, a poorer SEP (0.26) was obtained with the Catalyst 3D-QSAR model than with CoMFA or CoMSIA 3D-QSAR (Table 1). Predicted pIC_{50} values given by CoMFA, CoMSIA, and Catalyst 3D-QSAR models for the HDAC inhibitors training set are compared in Table 3 and Fig. S3.2. Apparently, Catalyst 3D-QSAR leads to more deviation in the predicted compared to measured pIC_{50} values for inhibitor b1, b7, b11, b13, b14, b19, b26, b27, b33, b40, b41, and b43 than either the CoMFA or CoMSIA 3D-QSAR models (Table 3). For this HDAC inhibitor training set, both the CoMFA and CoMSIA 3D-QSAR models give better predicted pIC_{50} values than those given by the Catalyst 3D-QSAR (Table 3).

The S and E contours on the structure of b1 detected by CoMFA and CoMSIA 3D-QSAR models for the HDAC inhibitor training set are compared in Fig. 2a–d. For these inhibitors, the S disfavored regions, represented by yellow, and S favored regions, represented by green contours, given by both the CoMFA and CoMSIA 3D-QSAR models are

Table 3 Actual (Act) and predicted (Pred) pIC_{50} values for the HDAC inhibitor training set derived from the best CoMFA, CoMSIA, and Catalyst 3D-QSAR models compared with those from the c-3D-QSAR model

Inhibitor#	Act pIC_{50}	CoMFA Pred pIC_{50}	CoMSIA (S+E + H) Pred pIC_{50}	Catalyst (hypothesis Hypo1) Pred pIC_{50}	c-3D-QSAR Pred pIC_{50}
b1	8.15	8.12	8.16	8.35	8.30
b3	7.82	7.82	7.82	7.80	7.82
b4	7.82	7.84	7.80	7.74	7.82
b6	7.52	7.53	7.55	7.57	7.47
b7	7.46	7.47	7.46	7.59	7.47
b9	7.22	7.23	7.22	7.22	7.26
b11	7.15	7.17	7.16	6.66	7.23
b13	7.12	7.14	7.13	6.89	7.13
b14	7.05	7.04	7.06	6.85	7.04
b19	6.96	6.86	6.79	6.70	6.87
b26	6.62	6.63	6.64	6.41	6.65
b27	6.62	6.61	6.61	6.49	6.58
b29	6.55	6.62	6.66	6.64	6.59
b32	6.38	6.42	6.39	6.68	6.42
b33	6.35	6.35	6.36	6.42	6.30
b38	6.23	6.21	6.25	6.24	6.17
b40	5.77	5.74	5.76	5.55	5.81
b41	5.70	5.68	5.70	6.09	5.70
b43	5.00	5.00	4.99	5.60	4.95

	CoMFA	CoMSIA
Steric	 a	 b
Electrostatic	 c	 d
	Catalyst	CoMSIA
Hydrophobic and Others	 e	 f
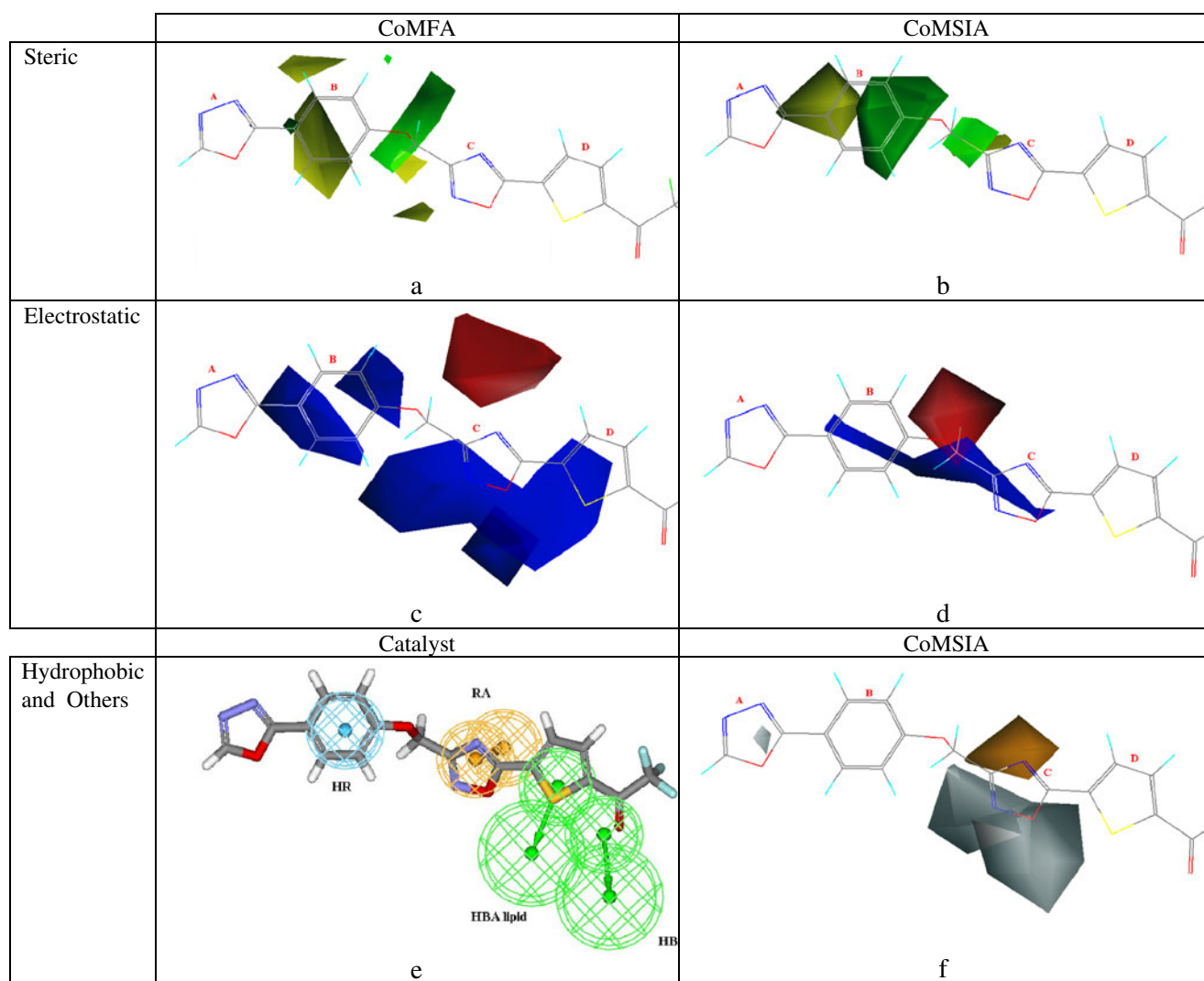
Fig. 2 a–f Pharmacophore features detected by the best CoMFA , CoMSIA, and Catalyst 3D-QSAR models built from the HDAC inhibitor training set (see legend of Fig. 1 for contour colors).The structure of **b1** is mapped onto the steric pharmacophore features of each model. **a** CoMFA steric contours of the best CoMFA 3D-QSAR model. **b** CoMSIA steric contours of the best CoMSIA 3D-QSAR model. **c** CoMFA electrostatic contours of the best CoMFA model. **d** CoMSIA electrostatic contours of the best CoMSIA model. **e** The structure of **b1** mapped onto the Catalyst 3D-QSAR pharmacophore features built from the HDAC inhibitor training set. Pharmacophore features: *cyan sphere* HR, *orange sphere* RA, *green sphere* HBA features. **f** CoMSIA H contours (*orange contours* hydrophobic favored regions, *white contours* hydrophobic disfavored regions) of the best CoMSIA 3D-QSAR model. The structure of **b1** is mapped onto the CoMSIA H pharmacophore features

almost the same (Fig. 2a,b). A similar phenomenon is observed by comparing the E features detected by both the CoMFA and CoMSIA 3D-QSAR models. The negative charge favored regions between rings B and C, depicted by red contours by both 3D-QSAR models, are exactly the same (Fig. 2c,d). However, more favored regions for positive charges along rings B, C, and D, represented with blue contours, are detected by CoMFA than by the CoMSIA 3D-QSAR model (Fig. 2c,d). The H features detected by both the CoMSIA and Catalyst 3D-QSAR models are compared in Fig. 2e,f. While the CoMSIA 3D-QSAR model detects a favored region for H feature and

marks it with orange contours near ring C, it also assigns the regions near ring C as disfavored regions for H feature and labels them with white contours (Fig. 2f). However, both regions near ring C and D are detected by the Catalyst 3D-QSAR model as favored regions for HBA (represented by green spheres) and RA (represented by orange spheres) features (Fig. 2e). The Catalyst 3D-QSAR model also identifies a HR feature represented by cyan spheres for ring B (Fig. 2e). Therefore, the hydrophobic features provided by the Catalyst 3D-QSAR can be used to supplement those given by the CoMSIA 3D-QSAR model for this inhibitor set.

Table 4 Actual (Act) and predicted (Pred) pIC$_{50}$ values for the FTase inhibitor training set derived from the best CoMFA, CoMSIA, and Catalyst 3D-QSAR models compared with those from the c-3D-QSAR model

Inhibitor #	Act pIC$_{50}$	CoMFA Pred pIC$_{50}$	CoMSIA (S+H+E) Pred pIC$_{50}$	Catalyst (hypothesis Hypo1) Pred pIC$_{50}$	c-3D-QSAR Pred pIC$_{50}$
c-1	9.80	9.80	9.72	9.74	9.82
c-4	9.70	9.64	9.54	9.14	9.57
c-6	9.40	9.33	9.35	9.15	9.34
c-7	9.37	9.40	9.48	9.32	9.54
c-8	9.35	9.40	9.32	9.05	9.28
c-11	9.24	9.25	9.26	9.21	9.17
c-15	9.12	8.96	8.94	9.12	9.13
c-16	9.09	9.16	9.19	9.14	9.10
c-17	9.08	9.15	9.09	9.00	9.18
c-21	9.01	8.91	8.95	8.89	8.94
c-23	9.00	8.94	8.97	9.00	8.97
c-24	8.77	8.78	9.01	8.72	8.90
c-26	8.72	8.79	8.78	9.17	8.72
c-28	8.70	8.76	8.87	8.85	8.89
c-30	8.09	8.09	8.10	8.15	8.16
c-32	7.92	7.95	7.91	7.55	7.82
c-34	7.24	7.27	7.12	7.34	7.26
c-35	7.17	7.22	7.21	7.64	7.17
d-2	9.46	9.28	9.28	9.08	9.22
d-4	9.80	9.28	9.24	8.92	9.19
d-5	9.16	9.19	9.15	9.04	9.19
d-7	9.11	9.10	9.08	9.15	9.07
d-11	9.06	9.10	9.16	9.21	9.22
d-12	9.06	9.15	9.03	9.00	9.02
d-13	9.04	8.94	9.01	8.85	8.94
d-14	9.03	9.07	9.02	8.85	9.07
d-15	9.02	8.99	8.98	9.08	8.90
d-17	9.01	8.99	8.98	8.44	8.80
d-19	8.96	8.91	9.09	8.96	9.07
d-21	8.96	8.90	9.12	9.11	9.18
d-22	8.96	8.94	9.01	9.11	8.97
d-23	8.96	8.98	8.94	9.12	8.97
d-25	8.89	8.73	8.66	9.04	8.73
d-26	8.80	8.82	8.79	8.96	8.83
d-28	8.70	8.88	8.88	8.92	8.90
d-29	8.70	8.84	8.68	9.00	8.65
d-30	8.66	8.70	8.64	8.19	8.69
d-32	8.40	8.40	8.39	9.00	8.45
d-35	8.28	8.10	8.24	8.46	8.28
d-36	8.05	8.10	8.11	8.60	8.11
d-37	7.96	7.92	8.02	7.96	7.94
d-38	7.89	7.87	7.92	7.96	7.85
d-40	7.46	7.43	7.44	7.57	7.50

Fig. 3 a–f Pharmacophore features detected by the best CoMFA , CoMSIA, and Catalyst 3D-QSAR models built from the FTase inhibitor training set (see legend of Fig. 1 for the definition of contour colors). The structure of **c-1** is mapped onto the pharmacophore features of each model. **a** CoMFA S contours of the best CoMFA 3D-QSAR model. **b** CoMSIA S contours of the best CoMSIA 3D-QSAR model. **c** CoMFA E contours of the best CoMFA 3D-QSAR model. **d** CoMSIA E contours of the best CoMSIA model. **e** Structure of **c-1** is mapped onto the Catalyst 3D-QSAR pharmacophore features built from the FTase inhibitor training set. Pharmacophore features: *cyan sphere* HR, *light blue sphere* Y, *orange sphere* RA features. **f** CoMSIA H contours (*orange contours* hydrophobic favored regions, *white contours* hydrophobic disfavored regions) of the best CoMSIA 3D-QSAR model. The structure of **c-1** is mapped onto the CoMSIA H pharmacophore features

Table 5 Summary of the normalized coefficient (NC) and fractional contribution (FC) computed by the SYBYL PLS module for each CoMFA, CoMSIA, and Catalyst 3D-QSAR and the c-3D-QSAR model built from each training set

| | BACE-1 inhibitors | | | | | | HDAC inhibitors | | | | | | FTase inhibitors | | | | | |
| | CoMFA | | CoMSIA | | c-3D-QSAR | | CoMFA | | CoMSIA | | c-3D-QSAR | | CoMFA | | CoMSIA | | c-3D-QSAR | |
	NC	FC	NC	FC	NC	FC	NC	FC	NC	FC	NC	FC	NC	FC	NC	FC	NC	FC
CoMFA																		
S	1.39	0.59			0.14	0.06	1.29	0.45			0.35	0.09	2.33	0.54			0.51	0.10
E	0.78	0.41			0.08	0.04	1.58	0.55			0.38	0.10	1.98	0.46			0.44	0.09
CoMSIA																		
S			0.32	0.15	0.05	0.02			0.33	0.12	0.11	0.03			0.60	0.16	0.16	0.03
E			0.72	0.34	0.11	0.05			1.47	0.54	0.32	0.09			1.99	0.54	0.69	0.13
H			0.64	0.31	0.10	0.04			0.93	0.34	0.28	0.08			1.10	0.30	0.50	0.10
A			0.41	0.20	0.06	0.03												
Catalyst																		
HBA1					0.22	0.10					0.31	0.08						
HBA2											0.55	0.15						
HY					0.65	0.29											0.59	0.11
HL					0.45	0.20												
HR1					0.19	0.08					0.79	0.21					0.93	0.18
HR2					0.22	0.10												
HR3																		
RA1											0.68	0.18					0.66	0.13
RA2																	0.69	0.13

3D-QSAR models of FTase inhibitors

The crystal structure 1NI1 [44] reveals that inhibitor c-19 binds with the FTase active site formed by a deep hydrophobic cleft between the α- and β-subunits (data not shown here). In fact, most of these FTase inhibitors (c- and d- series) are characterized by an interaction between the upper imidazole ring (ring B) and a zinc ion in the active site (Table S1.3) [44, 45]. However, a wide range of substituents have been placed at the para positions of the lower cyanophenyl ring (ring A) to increase the potency of these inhibitors (Table S1.3) [45]. The RMSD of the GOLD docked conformation computed for c-19 is 0.37 Å. The best CoMFA 3D-QSAR model obtained for the aligned FTase inhibitor training set was: $q_{loo}^2 = 0.66$, $q^2 = 0.66$, $r^2 = 0.98$, SEP = 0.09, and F = 375 (Table 1). Further, the interaction of these inhibitors with the FTase active site is best described by stepwise CoMSIA S+H+E field indexes (Table S2.3). These give the following statistics: $q_{loo}^2 = 0.69$, $q^2 = 0.67$, $r^2 = 0.98$, SEP = 0.102, and F = 285 (Table 1) for the best CoMSIA 3D-QSAR model obtained. The Catalyst 3D-QSAR model was built from the top hypothesis using the stepwise CoMSIA features and validated by the catScramble module. The cost differences computed between null and total 1 and that between null and fixed 1 were 64.3 and 87.9, respectively (Table S3.3). The overall statistics on the predicted pIC_{50} values for the FTase inhibitor training set given by the Catalyst 3D-QSAR model are: $q_{loo}^2 = 0.80$, $q^2 = 0.80$, $r^2 = 0.90$, SEP = 0.26, and F = 180 (Table 1). Measured versus predicted pIC_{50} values given by CoMFA, CoMSIA, and Catalyst 3D-QSAR models for the FTase inhibitor training set are compared in Table 4 and Fig. S3.3. However, more deviation was found in predicted compared to measured pIC_{50} values for c-4, c-6, c-8, c-26, c-28, c-32, c-34, c-35, d-2, d-4, d-17, d-29, d-30, d-32, d-35, d-36, and d-40 with Catalyst than with either the CoMFA or CoMSIA 3D-QSAR models (Table 4). For this FTase inhibitor training set, the pIC_{50} values predicted for c-24, c-28, c-34, d-4, and d-29 given

Fig. 5 pIC_{50} values of training and test set HDAC inhibitors predicted (Pred) by the c-3D-QSAR model plotted against actual (Act) pIC_{50} (Tables 3 and 7)

by CoMSIA 3D-QSAR or for d-4 and d-35 given by the CoMFA 3D-QSAR model deviated only slightly from the measured values (Table 4).

The S features of the structure of c-1 from the FTase inhibitor training set detected by both CoMFA and CoMSIA 3D-QSAR models were very similar (Fig. 3a,b). Regions near the naphthyl group on ring A or around ring A were identified by both CoMFA and CoMSIA 3D-QSAR models as favored or disfavored regions for S feature, as represented with green or yellow contours, respectively (Fig. 3a,b). However, some differences between the E features detected by CoMFA and CoMSIA 3D-QSAR models were observed as shown in Fig. 3c,d. First, regions near ring A or around amino acid residues D359β and W106β were identified as regions favored for positive charges (represented by blue contours) more extensively by CoMFA than by the CoMSIA 3D-QSAR model (Fig. 3c,d). Next, a region near the naphthyl group on ring A is assigned by CoMSIA 3D-QSAR as a region favored for positive charges (represented by blue contours) while the CoMFA 3D-QSAR model shows this as a region favored for negative charges (red contours in Fig. 3c,d). However, the region near the upper imidazole ring or zinc ion is unanimously assigned by both CoMFA and CoMSIA 3D-QSAR models as a region favored for negative charges (red contours in Fig. 3c,d). The H features detected by the

Fig. 4 pIC_{50} values of training and test set BACE-1 inhibitors predicted (Pred) by the c-3D-QSAR model plotted against actual (Act) pIC_{50} (Tables 2 and 6)
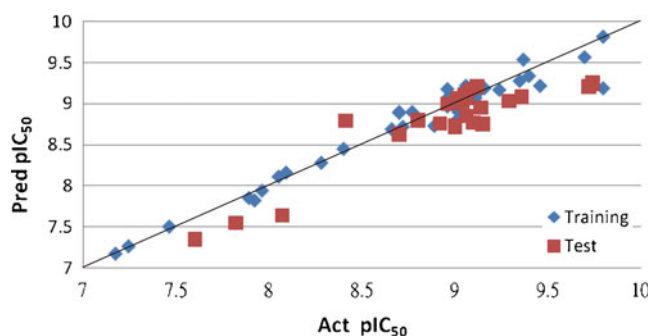
Fig. 6 pIC_{50} values of training and test set FTase inhibitors predicted (Pred) by the c-3D-QSAR model plotted against actual (Act) pIC_{50} (Tables 4 and 8)

Table 6 Actual (Act) and predicted (Pred) pIC_{50} values for the test set BACE-1 inhibitors by the best CoMFA, CoMSIA, and Catalyst 3D-QSAR models compared with those of the c-3D-QSAR model

Inhibitor #	Act pIC_{50}	CoMFA Pred pIC_{50}	CoMSIA (S+E+H+A) Pred pIC_{50}	Catalyst (hypothesis Hypo1) Pred pIC_{50}	c-3D-QSAR Pred pIC_{50}
a1	9.00	8.34	8.69	8.92	8.58
a2	8.70	8.36	8.25	8.55	8.68
a8	8.52	7.69	7.61	8.24	8.13
a9	8.40	8.34	8.28	8.43	8.50
a14	8.40	7.70	7.71	8.40	8.17
a15	8.40	8.11	7.69	8.44	8.21
a16	8.30	8.53	8.14	8.38	8.51
a19	8.10	8.00	8.09	8.15	7.98
a23	7.89	7.63	7.81	7.92	7.82
a25	7.82	7.69	8.15	7.92	7.81
a27	7.57	7.95	8.11	7.68	7.75
a32	5.96	6.15	5.89	5.96	6.50
a35	5.30	6.44	6.37	5.33	5.37
a38	4.70	6.20	6.30	5.05	5.16

CoMSIA and Catalyst 3D-QSAR models for the FTase inhibitor training set are compared in Fig. 3e,f, respectively. While both CoMSIA and Catalyst 3D-QSAR models identify the naphthyl group on ring A as the favored region for H features, marking them with orange contours, or cyan (HR) or light blue (HY) spheres, respectively, the former does not give any assignment near ring A and B, unlike the latter, which assigns these two regions as favored regions for RA features and labels them with orange spheres (Fig. 3e,f). However, there are three disfavored regions for H features expressed with white contours detected by the CoMSIA but not by the Catalyst 3D-QSAR model (Fig. 3e,f).

Consensus 3D-QSAR analyses

For all three inhibitor series studied, we find that the consensus in pharmacophores identified is high if the corresponding structures are aligned correctly and similar pharmacophore features are chosen for building both CoMFA and CoMSIA 3D-QSAR models. Pharmacophores that are undetectable in both the CoMFA and CoMSIA 3D-QSAR models may be supplemented with those from the Catalyst 3D-QSAR via construction of a c-3D-QSAR model on the same structurally aligned inhibitor set. Table 5 compares the SYBYL PLS analysis results obtained for each c-3D-QSAR for each inhibitor series

Table 7 Actual (Act) and predicted (Pred) pIC_{50} values for the test set HDAC inhibitors by the best CoMFA, CoMSIA, and Catalyst 3D-QSAR models compared with those of the c-3D-QSAR model

Inhibitor #	Act pIC_{50}	CoMFA Pred pIC_{50}	CoMSIA (S+E+H) Pred pIC_{50}	Catalyst (hypothesis Hypo1) Pred pIC_{50}	c-3D-QSAR Pred pIC_{50}
b5	7.52	7.93	7.97	7.29	7.55
b8	7.22	7.70	8.03	7.06	7.31
b15	7.05	7.58	7.37	6.90	7.21
b16	7.02	6.80	6.85	6.47	6.66
b17	7.00	6.11	6.25	7.00	6.69
b18	7.00	6.64	6.72	6.64	6.84
b20	6.96	6.77	6.68	6.42	6.53
b21	6.92	7.07	6.61	6.87	6.85
b23	6.77	5.86	5.65	7.16	6.57
b24	6.70	6.35	6.23	6.33	6.22
b25	6.66	6.50	6.35	6.85	6.71
b28	6.57	6.21	6.24	6.40	6.09
b30	6.52	5.56	5.84	6.32	5.95
b34	6.32	5.34	5.35	5.94	5.36
b37	6.28	5.33	5.49	6.04	5.45

with those obtained for each CoMFA and CoMSIA 3D-QSAR model. Based on the normalized coefficients (NC) computed, SYBYL PLS analyses assign a fractional contribution (FC) for each pharmacophore feature selected for building each CoMFA or CoMSIA 3D-QSAR as well as the c-3D-QSAR model. Since only S and E features are used by CoMFA, the FC by each feature to each CoMFA 3D-QSAR model built for each inhibitor series computed is nearly the same (Table 5). However, for the CoMSIA 3D-QSAR models, the computed FC varies from 0.15 (S feature) to 0.34 (E feature) for the BACE-1, 0.12 (S feature) to 0.54 (E feature) for the HDAC, and 0.16 (S feature) to 0.54 (E feature) for the FTase inhibitor series, respectively (Table 5). These results indicate that, while the E feature remains somewhat the same, the S feature of the original CoMFA 3D-QSAR is split rationally into either the S+H+A or S+H features of the CoMSIA 3D-QSAR models (Table 5). As selected by the stepwise CoMSIA runs, most of the Catalyst pharmacophore features used to build the c-3D-QSAR models are hydrophobic-related features. The computed FC of total CoMFA, CoMSIA, and Catalyst pharmacophore features

selected to build each c-3D-QSAR model are 0.10, 0.14, and 0.77 for the BACE-1 inhibitor series; 0.19, 0.20, and 0.61 for the HDAC inhibitor series; and 0.19, 0.26, and 0.55 for the FTase inhibitor series (Table 5). This clearly shows that the contribution of Catalyst hydrophobic features is far more significant than that of the CoMFA and CoMSIA features used in building each c-3D-QSAR model. Note that the computed FC by E feature is reduced significantly from 0.34, 0.54, and 0.54 for a CoMSIA 3D-QSAR to 0.05, 0.09, and 0.13 for the c-3D-QSAR model built for each inhibitor series (Table 5).

As judged by q_{loo}^2, q^2, r^2, SEP, and F computed for each training set, each c-3D-QSAR apparently outperforms each 3D-QSAR model built for each inhibitor training set (Table 1). Although somewhat better q_{loo}^2 and q^2 values are obtained by Catalyst 3D-QSAR than by the c-3D-QSAR model built for the BACE-1 inhibitor series, the SEP computed by the former is somewhat poorer than that by the latter (Table 1). Similar phenomena are observed for both the HDAC and FTase inhibitor series, i.e., much worse SEP while somewhat better q_{loo}^2 and q^2 values are obtained by Catalyst 3D-QSAR than by c-3D-QSAR

		CoMFA	CoMSIA (S+H+E)	Catalyst (hypothesis Hypo1)	c-3D-QSAR
Inhibitor #	Act pIC_{50}	Pred pIC_{50}	Pred pIC_{50}	Pred pIC_{50}	Pred pIC_{50}
c-2	9.74	9.72	9.40	9.09	9.24
c-3	9.74	9.24	9.17	9.18	9.26
c-10	9.29	9.44	9.10	9.08	9.03
c-13	9.15	9.06	8.84	8.82	8.75
c-14	9.14	8.43	8.71	9.00	8.95
c-18	9.06	9.26	8.67	9.06	8.85
c-19	9.05	8.94	9.18	9.17	9.11
c-20	9.04	9.07	9.00	9.15	9.00
c-22	9.00	8.69	8.80	9.43	8.72
c-27	8.70	9.14	8.94	8.33	8.63
c-29	8.41	8.83	8.74	9.00	8.79
c-31	8.07	8.01	7.62	7.38	7.64
c-33	7.60	7.94	7.84	7.27	7.35
d-1	9.72	9.39	9.13	9.15	9.21
d-3	9.36	9.11	9.03	9.17	9.08
d-6	9.12	9.28	9.06	8.80	9.21
d-8	9.10	8.59	8.83	9.09	8.77
d-9	9.08	8.97	9.13	9.15	9.17
d-10	9.06	8.99	9.16	9.14	9.07
d-16	9.01	9.21	8.90	9.06	9.06
d-18	9.00	8.89	8.90	8.77	8.73
d-20	8.96	8.78	8.94	8.96	9.00
d-24	8.92	8.72	8.73	8.80	8.76
d-27	8.80	8.94	8.59	9.14	8.80
d-39	7.82	8.11	8.12	7.43	7.55

Table 8 Actual (Act) and predicted (Pred) pIC_{50} values for the test set FTase inhibitors by the best CoMFA, CoMSIA, and Catalyst 3D-QSAR models compared with those of the c-3D-QSAR model

modeling (Table 1). With the exception of the results above, most of the q_{loo}^2, q^2, r^2, and SEP values computed by c-3D-QSAR are superior to those from CoMFA or CoMSIA 3D-QSAR models built for each inhibitor training set (Table 1). The predicted pIC_{50} values given by each c-3D-QSAR for some less active inhibitors in each training set are also better than those from CoMFA, CoMSIA, and Catalyst 3D-QSAR models as shown in Tables 2, 3, and 4, respectively. For examples, the predicted pIC_{50} for a30 and a33 (BACE-1 inhibitors), b29 and b41 (HDAC inhibitors), and c-34 and c-35 (FTase inhibitors) given by c-3D-QSAR are apparently better than those from each 3D-QSAR model (Tables 2, 3, 4; Figs. 4, 5, 6). The c-3D-QSAR also outperforms the individual CoMFA, CoMSIA, and Catalyst 3D-QSAR models in predicting pIC_{50} for inhibitors of each test set as shown in Tables 1, 6, 7, 8 and Figs. 4–6. With the exception of the BACE-1 test set, the r^2 values obtained with c-3D-QSAR are clearly better than those from each CoMFA, CoMSIA, and Catalyst 3D-QSAR model built for each test set (Table 1). The Catalyst prediction on the BACE-1 test set is exceptionally good, since the r^2 computed for both training and test sets are very close (Table 1). Moreover, the pIC_{50} of some less active inhibitors of each test set are predicted more accurately by each c-3D-QSAR than by individual CoMFA, CoMSIA, and Catalyst 3D-QSAR models as shown in Tables 6, 7, and 8. For example, the pIC_{50} values of a25, a35, and a38 of the BACE-1 series, and those of b5, b8, b15, b18, and b25 of the HDAC series are more accurately predicted by the corresponding c-3D-QSAR than by each of the CoMFA, CoMSIA, and Catalyst 3D-QSAR models (Tables 6, 7).

Conclusions

In this report, we have presented a simple method for building c-3D-QSAR models by using the entire CoMFA and CoMSIA results and some Catalyst pharmacophore features judiciously selected from prior stepwise CoMSIA runs. Our aim was to improve prediction accuracy on some of the less active compounds often encountered in building CoMFA, CoMSIA, or Catalyst 3D-QSAR models. The primary requirement for building a c-3D-QSAR model is that ligand structures must be generated correctly and aligned, which is similar to the requirements for building a 3D-QSAR model. However, how to choose the correct pharmacophore features to use in the CoMSIA or Catalyst method is the most critical parameter in building an accurate 3D-QSAR model, since the true pharmacophore features of a receptor active site are usually unknown. Fortunately, delicate statistical methods, e.g., PLS and others, have been developed that can be used to rationally

train a dataset by employing the desired pharmacophore features. Here, we have shown that the performance of CoMFA 3D-QSAR models built from simple pharmacophore features can be better than that of some CoMSIA or Catalyst 3D-QSAR models built from more versatile pharmacophore features. The key point is that there is plenty room to manipulate pharmacophore features in order to build more accurate 3D-QSAR models. Choosing some preliminary stepwise CoMSIA runs, we found that most of the Catalyst pharmacophore features used are hydrophobic and complementary to those used in building CoMFA and CoMSIA 3D-QSAR models. Unfortunately, these Catalyst pharmacophore features cannot be blended automatically with the CoMFA and CoMSIA ones used by the SYBYL PLS module for making more genuine pharmacophores for the c-3D-QSAR models. Nevertheless, we have shown that the problem of incorrect prediction often encountered for some less active compounds by a CoMFA, CoMSIA, and Catalyst 3D-QSAR was improved substantially by all the c-3D-QSAR models built here.

References

1. Croog SH, Burleson JA, Sudilovsky A, Baume RM (2006) Spouse caregivers of Alzheimer patients: problem responses to caregiver burden. Aging Ment Health 10:87–100
2. Hardy J, Selkoe DJ (2002) Thyamyloid hypothesis of Alzheimer's disease: progress and problems on the road to the therapeutics. Science 297:353–356
3. De Strooper BD, Köning G (1999) Alzheimer's disease a firm base for drug development. Nature 402:471–472
4. Plaques and Tangles: The Hallmarks of AD, A. D. E. A. R Center, National Institute on Aging, http://www.nia.nih.gov/Alzheimers/Publications/UnravelingTheMystery/Part1/Hallmarks.htm
5. Cai H, Wang Y, McCarthy D, Wen H, Borchelt DR, Price DL, Wong PC (2001) BACE1 is the major β-secretase for generation of Aβpeptides by neurons. Nat Neurosci 4:233–234
6. Iserloh U, Pan J, Stamford AW, Kennedy ME, Zhang Q, Zhang L, Parker EM, McHugh NA, Favreau L, Strickland C, Voigt J (2008) Discovery of an orally efficacious 4-phenoxypyrrolidine-based BACE-1 inhibitors. Bioorg Med Chem Lett 18:418–422
7. Lindsley SR, Moore KP, Rajapakse HA, Seinick HG, Young MB, Zhu H, Munshi S, Kuo L, McGaughey GB, Colussi D, Crouthamel MC, Lai MT, Pietrak B, Price EA, Sankaranarayanan S, Simon AJ, Seabrook GR, Hazuda DJ, Pudvah NT, Hochman JH, Graham SL, Vacca JP, Nantermet PG (2007) Design, Synthesis, and SAR of macrocyclic tertitary carbinamine BACE-1 inhibitors. Bioorg Med Chem Lett 17:4057–4061
8. Stachel SJ, Colourn CA, Steele TG, Crouthamel MC, Pietrak BL, Lai MT, Holloway MK, Munshi SK, Graham SL, Vacca JP (2006) Conformationally based P3 amide replacements of beta-secretase inhibitors. Bioorg Med Chem Lett 16:641–644

9. Limongelli V, Marinelli L, Cosconati S, Braun HA, Schmidt B, Novellino E (2007) Ensemble-docking approach on BACE-1: pharmacophore perception and guidelines for drug design. Chem Med Chem 2:667–678

10. Gerschwinder S, Olsson LL, Albert JS, Deinum J, Edwards PD, de Beer T, Folmer RH (2007) Discovery of a novel warhead against beta-secretase through fragment-based lead generation. J Med Chem 50:5903–5911

11. Yang W, Fucini RV, Fahr BT, Randal M, Lind KE, Lam MB, Lu W, Lu Y, Cary DR, Ramanowski MJ, Colussi D, Pietrak B, Allison TJ, Munshi SK, Penny DM, Pham P, Sun J, Thomas AE, Wilkinson JM, Jacobs JW, McDowell RS, Ballinger MD (2009) Fragment-based discovery of nonpeptidic BACE-1 inhibitors using tethering. Biochemistry 48:4488–4496

12. Cramer RD III, Patterson DE, Bunce JS (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967

13. 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

14. Pandey A, Mungalpara J, Mohan CG (2010) Comparative molecular field analysis and comparative molecular similarity indices analysis of hydroxyethylamine derivatives as selective human BACE-1 inhibitor. Mol Divers 14:39–49

15. Pazin MJ, Kadonaga JT (1997) What's up and down with histone deacetylation and transcription? Cell 89:325–328

16. Stemer DE, Berger SL (2000) Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev 64:435–439

17. Konstantinopoulos PA, Karamouzis MV, Papavassiliou AG (2007) Focus on acetylation: the role of histone deacetylase inhibitors in cancer therapy and beyond. Expert Opin Investig Drugs 16:569–571

18. Yosida M, Kijima M, Akita M, Beppu T (1990) Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin. J Biol Chem 265:17174–17179

19. Richon VM, Emiliani S, Verdin E, Webb Y, Breslow R, Rifkind RA, Marks PA (1998) A class of hybrid polar inducers of transformed cell differentiation inhibits histone deacetylase. Proc Natl Acad Sci USA 95:3003–3007

20. Plumb JA, Finn PW, Williams RJ, Bandara MJ, Romero MR, Watkins CJ, La Thangue NB, Brown R (2003) Pharmacodynamic response and inhibition of growth of human tumor xenografts by the novel histone deacetylase inhibitor PXD101. Mol Cancer Ther 2:721–728

21. Suzuki T, Ando T, Tsuchiya K, Fukazawa N, Saito A, Mariko Y, Yamashita T, Nakanishi O (1999) Synthesis and histone deacetylase inhibitory activity of new benzamide derivatives. J Med Chem 42:3001–3003

22. Leoni F, Zaliani A, Bertolini G, Porro G, Pagani P, Pozzi P, Dona G, Fossati G, Sozzani S, Azam T, Bufler P, Fantuzzi G, Goncharov I, Kim SH, Pomerantz BJ, Reznikov LL, Siegmund B, Dinarello CA, Mascagni P (2002) The antitumor histone deacetylase inhibitor suberoylanilide hydroxamic acid exhibits anti-inflammatory properties via suppression of cytokines. Proc Natl Acad Sci USA 99:2995–3000

23. De Ruijter AJ, van Gennip AH, Caron HN, Kemp S, van Kuilenburg AB (2003) Histone deacetylase (HDACs): characterization of the classical HDAC family. Biochem J 370:737–749

24. Grant S, Easley C, Kirkpatrick P (2007) Vorinostat Nat. Rev Drug Discov 6:21–22

25. Ryan QC, Headlee D, Acharya M, Sparreboom A, Trepel JB, Ye J, Figg WD, Hwang K, Chung EJ, Murgo A, Melillo G, Elsayed Y, Monga M, Kalnitskiy M, Zwiebel J, Sausville EA (2005) Phase 1 and pharmacokinetic study of MS-275, a histone deacetylase inhibitors in patients with advanced and refractory solid tumors or lymphoma. J Clin Oncol 23:3912–3922

26. Byrd JC, Marcucci G, Parthun MR, Xiao JJ, Klisovic RB, Moran M, Lin TS, Liu S, Sklenar AR, Davis ME, Lucas DM, Fisher B, Shank R, Tejaswi SL, Binkley P, Wright J, Chan KK, Grever MR (2005) A phase 1 and pharmacodynamic study of depsipeptide (FK228) in chronic lymphocytic leukemia and acute myeloid leukemia. Blood 105:959–967

27. Wang DF, Wiest OG, Helquist P, Lan-Hargest HY, Wiech NL (2004) QSAR studies of PC-3 cell line inhibition activity of TSA and SAHA-like hydroxamic acids. Bioorg Med Chem Lett 14:707–711

28. Xie A, Liao C, Li Z, Ning Z, Hu W, Lu X, Shi L, Zhou J (2004) Quantitative structure-activity relationship study of histone deacetylase inhibitors. Curr Med Chem Anticancer Agents 4:273–299

29. Juvale DC, Kulkarni VV, Deokar HS, Wagh NK, Padhye SB, Kulkarni VM (2006) 3D-QSAR of histone deacetylase inhibitors: hydroxamate analogues. Org Biomol Chem 4:2858–2868

30. Bos JL (1989) Ras oncogenes in human cancer: a review. Cancer Res 49:4682–4689

31. Gibbs JB (1991) Ras C-terminal processing enzymes—new drug targets? Cell 65:1–4

32. Willumsen BM, Norris K, Papageorge AG, Hubbert NL, Lowy DR (1984) Harvey murine sarcoma virus p21 ras protein: biological and biochemical significance of the cysteine nearest the carboxy terminus. EMBO J 3:2581–2585

33. Casey PJ, Solski PA, Der CJ, Buss JE (1989) p21 ras is modified by a farnesyl isoprenoid. Proc Natl Acad Sci USA 86:8323–8327

34. Roberts PJ, Der CJ (2007) Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. Oncogene 26:3291–3310

35. Chen WJ, Andres DA, Goldstein JL, Brown MS (1991) Cloning and expression of a cDNA encoding the alpha subunit of rat p21ras protein farnesyltransferase. Proc Natl Acad Sci USA 88:11368–11372

36. Chen WJ, Andres DA, Goldstein JL, Russell DW, Brown MS (1991) cDNA cloning and expression of the peptide-binding beta subunit of rat p21ras farnesyltransferase, the counterpart of yeast DPR1/RAM1. Cell 66:327–334

37. Puntambekar DS, Giridhar R, Yadav MR (2008) Insights into the structural requirements of farnesyltransferase inhibitors as potential anti-tumor agents based on 3D-QSAR CoMFA and CoMSIA models. Eur J Med Chem 43:142–154

38. Leonard DM (1997) Ras farnesyltransferase: a new therapeutic target. J Med Chem 40:2971–2990

39. Perrin D, Halazy S, Hill B (1997) Inhibitors of Ras farnesyl-transferase: tomorrows' anticancer agents? Bull Cancer 84:635–642

40. Gibbs JB, Graham SL, Hartman GD, Koblan KS, Kohl NE, Omer CA, Oliff A (1997) Farnesyltransferase inhibitors versus Ras inhibitors. Curr Opin Chem Biol 1:197–203

41. Qian Y, Sebti SM, Hamilton AD (1997) Farnesyltransferase as a target for anticancer drug design. Biopolymers 43:25–41

42. Sun J, Qian Y, Hamilton AD, Sebti SM (1998) Both farnesyl-transferase and geranylgeranyltransferase 1 inhibitors are required for inhibition of oncogenic K-Ras prenylation but each alone is sufficient to suppress human tumor growth in nude mouse xenografts. Oncogene 16:1467–1473

43. Hunt JT, Lee VG, Leftheris K, Seizinger B, Carboni J, Mabus J, Ricca C, Yan N, Manne V (1996) Potent cell active non-thiol tetrapeptide inhibitors of farnesyltransferase. J Med Chem 39:353–358

44. Tong Y, Lin NH, Wang L, Hasvold L, Wang W, Leonard N, Li T, Li Q, Cohen J, Gu WZ, Zhang H, Stoll V, Bauch J, Marsh K, Rosenberg SH, Sham HL (2003) Discovery of potent imidazole and cyanophenyl containing farnesyltransferase inhibitors with improved oral bioavailability. Bioorg Med Chem Lett 13:1571–1574

45. Wang L, Wang GT, Wang X, Tong Y, Sullivan G, Park D, Leonard NM, Li Q, Cohen J, Gu WZ, Zhang H, Bauch JL, Jakob CG, Hutchins CW, Stoll VS, Marsh K, Rosenberg SH, Sham HL, Lin NH (2004) Design, synthesis, and biological activity of 4-[(4-cyano-2-arylbenzyloxy0-(3-methyl-3 H-imidazol-4-yl) methyl]benzonitriles as potent and selective farnesyltransferase inhibitors. J Med Chem 47:612–626

46. Xie A, Odde S, Prasanna S, Doerksen RJ (2009) Imidazole-containing farnesyltransferase inhibitors: 3D quantitative structure-activity relationships and molecular docking. J Comput Aided Mol Des 23:431–448

47. Equbal T, Silakari O, Rambabu G, Ravikumar M (2007) Pharmacophore mapping of diverse classes of farnesyltransferase inhibitors. Bioorg Med Chem Lett 17:1594–1600

48. Stahle L, Wold S (1986) On the use of some multivariate statistical methods in pharmacological research. J Pharmacol Meth 16:91–110

49. Catalyst 4.11, Accelrys Inc., San Diego, CA 2005, http://accelrys.com/.

50. Bharatham N, Bharatham K, Lee KW (2006) Pharmacophore identification and virtual screening for methionyl-tRNA synthetase inhibitors. J Mol Graph Model 25:813–823

51. Freskos JN, Fobian YM, Benson TE, Bienkowski MJ, Brown DL, Emmons TL, Heintz R, Laborde A, McDonald J, Mischke BV, Molyneaux JM, Moon JB, Mullins PB, Prince DB, Poddock DJ, Tomasselli AG, Winterrowd G (2007) Design of potent inhibitors of human β-secretase. Part 1. Bioorg Med Chem Lett 17:73–77

52. Freskos JN, Fobian YM, Benson TE, Moon JB, Bienkowski MJ, Brown DL, Emmons TL, Heintz R, Laborde A, McDonald JJ, Mischke BV, Molyneaux JM, Mullins PB, Prince B, Poddock DJ, Tomasselli AG, Winterrowd G (2007) Design of potent inhibitors of human β-secretase. Part 2. Bioorg Med Chem Lett 17:78–81

53. Wei HY, Tsai KC, Lin TH (2005) Modelling ligand-receptor interaction for some MHC class II HLA-DR4 peptide mimetic inhibitors using several molecular docking and 3D QSAR techniques. J Chem Inf Model 45:1343–1351

54. Yang L, Tan CH, Hsieh MJ, Wang J, Duan Y, Cieplak P, Caldwell J, Kollman PA, Luo R (2006) New-generation Amber united-atom force field. J Phys Chem B 110:13166–13176

55. SYBYL 8.0, Tripos Associates, St. Louis, MO

56. GOLD User Guide & Tutorials (2006) The Cambridge Crystallographic Data Centre. http://www.ccdc.cam.ac.uk/

57. Jakalian A, Jack DB, Bayly CI (2002) Fast efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem 23:1623–1641

58. Case DA, Darden TA, Cheatham TEIII, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2006) AMBER 9. University of California, San Francisco

59. Scarpelli R, Marco AD, Ferrigno F, Laufer R, Marcucci I, Muraglia E, Ontoria JM, Rowley M, Serafini S, Steinkühler C, Jones P (2008) Studies of the metabolic stability in cells of 5-(trifluoroacetyl)thiophene-2-carboxamides and identification of more stable class II histone deacetylase (HDAC) inhibitors. Bioorg Med Chem Lett 18:6078–6082

60. Muraglia E, Altamura S, Branca D, Cecchetti O, Ferrigno F, Orsale MV, Palumbi MC, Rowley M, Scarpelli R, Steinkühler C, Jones P (2008) 2-trifluoroacetylthiophene oxadiazoles as potent and selective class II human histone deacetylase inhibitors. Bioorg Med Chem Lett 18:6083–6087

61. Bottomley MJ, Lo Surdo P, Di Giovine P, Cirillo A, Scarpelli R, Ferrigno F, Jones P, Neddermann P, De Francesco R, Steinkühler C, Gallinari P, Carfi A (2008) Structural and functional analysis of the human Hdac 4 catalytic domain reveals a regulatory zinc-binding domain. J Biol Chem 283:26694–26704

ORIGINAL PAPER

Insights from ligand and structure based methods in virtual screening of selective Ni-peptide deformylase inhibitors

Ravi Shekar Ananthula · Muttineni Ravikumar · S. K. Mahmood · M. N. S. Pavan Kumar

Abstract In recent years, there has been a growing interest in developing bacterial peptide deformylase (PDF) inhibitors as novel antibiotics. The purpose of the study is to generate a three-dimensional (3D) pharmacophore model by using diverse PDF inhibitors which is useful for designing of potential antibiotics. Twenty one structurally diverse compounds were considered for the generation of quantitative pharmacophore model using HypoGen of Catalyst, further model was validated using 78 compounds. Pharmacophore model demonstrated the importance of two acceptors, one donor and one hydrophobic feature toward the biological activity. The inhibitors were also docked into the binding site of PDF to comprehend the structural insights of the active site. Combination of ligand and structure based methods were used to find the potential antibiotics.

Keywords Induced fit docking · Pharmacophore · Staphylococcus aureus · Zn metal binding proteins

R. S. Ananthula (✉) · M. N. S. P. Kumar
BioCampus, GVKBIO S-1, Phase-1,
Technocrats Industrial Estate,
Balanagar,
Hyderabad, Andhra Pradesh 500037, India
e-mail: a.ravishekar@gmail.com

R. S. Ananthula · M. Ravikumar · S. K. Mahmood
Bioinformatics Division,
Environmental Microbial Laboratory Department of Botany,
Osmania University,
Hyderabad, Andhra Pradesh 500007, India

Introduction

Bacterial resistance to many of the existing antibiotics is a growing health concern [1–3]. Over the past decade, there has been an alarming increase in the prevalence of multiresistant phenotypes among pathogenic bacteria and hence there is an urgent need to identify novel antibiotics with unexploited modes of action. In recent years, there has been a growing interest in developing bacterial Ni-peptide deformylase (PDF) inhibitors as novel antibiotics. PDF (EC 3.5.1.88) is an essential bacterial metallo-enzyme responsible for the removal of the N-terminal formyl group from methionine residues following protein synthesis [4, 5] and considered as an attractive target for antibacterial chemotherapy [6, 7]. To date several classes of inhibitors have already been investigated, [6–10]. Among them actinonin is the available drug derived from natural products, including inhibiting peptide deformylase (PDF) [11–13]. Most of the compounds with sufficient potency and antibacterial activity share a common structure, as exemplified the presence of Acetamide derivatives in the literature [14–19]. These compounds with pseudopeptidic backbone have low specificity for matrix metalloproteins (MMP) and poor metabolic stability associated with pseudopeptidic inhibitors, there is a need for the identification of novel and structurally diverse non-peptidic PDF inhibitors.

PDF is essential for both Gram-positive and Gram-negative bacteria, since deformylation is a necessary step to complete protein biosynthesis and maturation. Two bacterial PDF types, PDF1B and PDF2, have been distinguished on the basis of their functionality and presence/absence in the Gram positive/negative bacteria [13, 20]. PDF2 is found only in Gram-positive bacteria. Bacteria may have one or several functional genes encoding different types of PDF

[13, 20] for example, Gram-negative *Escherichia coli* have one PDF1B gene (def) only, whereas Gram-positive bacteria such as *Bacillus* spp. have two PDF genes: def, encoding PDF1B, and ykrB, encoding PDF2. The human PDF (mPDF) has been classified as a PDF1A [21]. The recent revelation of the existence of a PDF homologue (mitochondrial PDF or mPDF) in humans [7, 22, 23] raised major objections to the use of this otherwise very promising target [20, 24]. The aim of our study was to identify new compounds that would selectively inhibit.

Identification of a potent PDF inhibitor is no guarantee that it will be active against intact bacterial cells. The ability of an inhibitor to access the target (by penetrating the permeability barrier), and to maintain an effective concentration (in the face of efflux pumps and inactivation) is also critical. Fortunately many potent PDF inhibitors have also proved to be active against bacteria. Typically these compounds have contained hydroxamates or reverse hydroxamates as chelators. The antibacterial spectrum of PDF inhibitors published to date has been primarily gram positive, including staphylococci and streptococci, and there is no cross-resistance with existing antibiotic classes [8].

As staphylococcus Ni-peptide deformylase (PDF) is a metallo protein having zinc metal in the binding site. The available hydroxamate class of antibiotics is considered the most effective zinc binding group but they lack selectivity, especially they can interact with human physiologically important metallo enzymes. Hence there is an urgent need to identify selective inhibitors of PDF. In the present study we have generated the best predictive pharmacophore model using diverse inhibitors of PDF and performed docking experiments to elucidate the important interactions between PDF and its inhibitors responsible for biological activity and selectivity. Using these two models we have performed virtual screening and identified the potential inhibitors against PDF.

Materials and methods

Pharmacophore modeling correlates activities with the spatial arrangement of various chemical features in the molecules. For the pharmacophore modeling studies, a total of 99 Ni-peptide deformylase (PDF) inhibitors with activity data (IC$_{50}$) spanning over 5 orders of magnitude (from 2 to 100,000 nM) were selected from the literature [25–30]. Biological activities of these molecules were obtained from similar assay method. The activity was determined as inhibitory concentration of the compound against staphylococcus aureus peptide deformylase using formate dehydrogenase (FDH)-coupled assay. The total dataset was further divided into a training and test set. The training set was selected considering both the structural diversity and wide coverage of the activity. Training set included the most

active, several moderately active, and some inactive compounds in order to obtain critical information on pharmacophore requirements. The important aspect of this selection scheme was that each active compound would teach something new to the HypoGen module to help it uncover as much critical information as possible for predicting biological activity. The training set consisted of 21 compounds selected with the above criteria. To validate the pharmacophore model, the other 78 compounds were used as the test set (S1). The reported activities (IC$_{50}$) against PDF are classified as: highly active (<10 nM), moderately active (10–100 nM), and inactive (>100 nM). All the IC$_{50}$ values were determined using the same experimental/assay method.

The structures of all the compounds were built from fragments in Catalyst 4.10 [31–33]. A CHARMm like force field [34] in the Catalyst program was utilized to ascertain the energy-minimized conformations for each structure. Details of the pharmacophore development procedures have been described in the literature [35–37]. Initially, conformational models of all molecules for PDF datasets were generated using the 'best quality' conformational search option within the Catalyst ConFirm module using the 'Poling' algorithm [38]. A maximum of 250 conformations were generated for each compound to ensure maximum coverage in the conformational space within an energy threshold of 20.0 kcal mol^{-1} above the global energy minimum. Instead of using just the lowest energy conformation of each compound, all conformational models for each molecule in training set were used in Catalyst for pharmacophore hypothesis generation.

Pharmacophore generation and validation

From the conformers of the training set compounds and their experimentally determined inhibitory activities against PDF, 10 best pharmacophore (also called hypotheses in the program) models were generated using HypoGen module implemented in Catalyst 4.10 software. An initial analysis revealed that three chemical feature types, i.e., hydrogen-bond acceptor (HBA), hydrogen-bond donor (HBD), and hydrophobic (HY) features could effectively map all critical chemical features of all molecules in the training and test sets. These features were selected and used to build a series of hypotheses with the HypoGen module in Catalyst using default uncertainty value 3 (defined by Catalyst as the measured value being within three times higher or three times lower of the true value).

Indeed, Catalyst generates a chemical-feature-based model on the basis of the most active compounds. These compounds are determined by performing a simple calculation based on the activity and uncertainty. As a matter of fact, the activity of the most active compound is multiplied by the uncertainty to

establish a comparison number, 'A'. The activity of the next most active compound is divided by the uncertainty, and this result in 'B', which is then compared to A. If B is smaller than A, the compound is included in the most active set; if not, the procedure stops [39].

In hypothesis generation, the structure and activity correlations in the training set were rigorously examined. HypoGen identifies features that are common to the active compounds but excluded from the inactive compounds within conformation allowable regions of space. Additionally, it estimates the activity of each training set compound using regression parameters. The parameters are computed by the regression analysis using the relationship of geometric fit value versus the negative logarithm of activity. The greater the geometric fit, the greater the activity prediction of the compound. The fit function not only checks if the feature is mapped or not, but also contains a distance term, which measures the distance that separates the feature on the molecule from the centroid of the hypothesis feature. Both terms are used to calculate the

geometric fit value. The generated pharmacophore model should be statistically significant, should predict activity of the molecules accurately, and should identify active compound from a database. Therefore, the derived pharmacophore map was validated using (i) cost analysis, (ii) test set prediction, and (iii) enrichment factor.

Cost analysis

The HypoGen module in Catalyst performs two important theoretical cost calculations (represented in bit units) that determine the success of any pharmacophore hypothesis. One is the 'fixed cost' (also termed as ideal cost), which represents the simplest model that fits all data perfectly, and the second one is the 'null cost' (also termed as no correlation cost), which represents the highest cost of a pharmacophore with no features and estimates activity to be the average of the activity data of the training set molecules. A meaningful pharmacophore hypothesis may result when the difference between
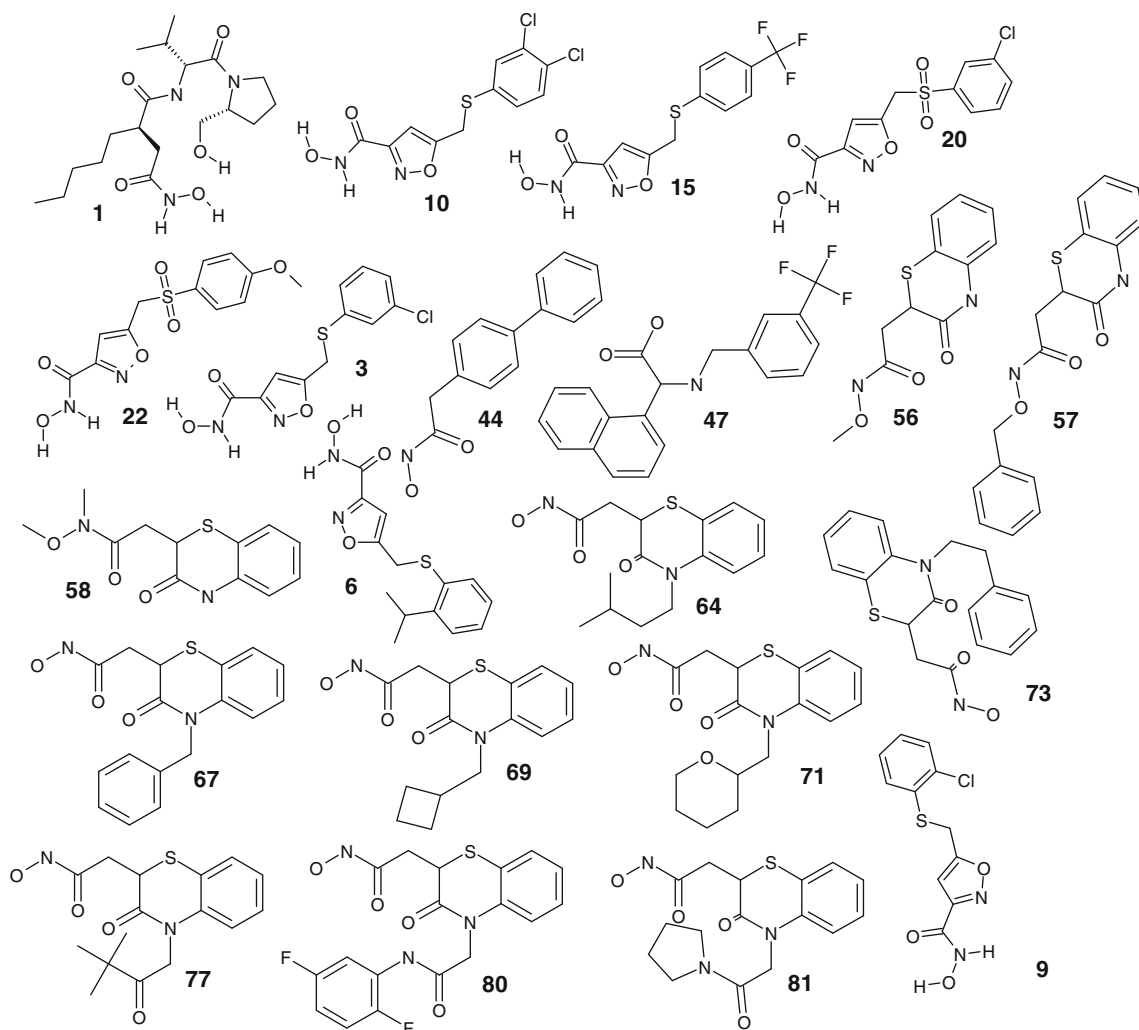
Chart 1 Structures of 21 training set molecules

null and fixed cost value is large; a value of 40–60 bits for a pharmacophore hypothesis may indicate that it has 75–90% probability of correlating the data (Catalyst 4.10 documentation).

The total cost (pharmacophore cost) of any pharmacophore hypothesis should be close to the fixed cost to provide any useful models. Two other parameters that also determine the quality of any pharmacophore hypothesis with possible predictive values are the configuration cost or entropy cost, which depends on the complexity of the pharmacophore hypothesis space and should have a value <17, and the error cost, which is dependent on the root mean square differences between the estimated and the actual activities of the training set molecules. The root mean square deviation (RMSD) represents the quality of the correlation between the estimated and the actual activity data. The best pharmacophore model has the highest cost difference, lowest RMSD, and best correlation coefficient.

Test set activity prediction

In addition to estimation of activity of training set molecules, the pharmacophore model should also esti-

mate the activity of new compounds. Therefore, a set of 78 PDF actives (Chart S1), which were not included in training set, was considered as a test set. These molecules are covering a wide range of IC50 activities spanning from 5 to 100000 nM.

Enrichment of database In the lead-discovery studies, the pharmacophore model should identify active leads against NPDs in the database screening. Therefore, an in-house database of 2 million molecules was spiked with 21 known inhibitors in order to validate whether the pharmacophore model could identify active compounds. This spiked database (containing 1525 molecules) was screened with the pharmacophore model and the enrichment factor (E) [35] was calculated using:

$$E = \text{Ha}/\text{Ht} \div A/D, \tag{1}$$

where Ht=the number of hits retrieved, Ha=the number of actives in the hit list, A=the number of active molecules present in the database, and D=the total number of molecules in the database.

Table 1 Experimental and 'Hypo 1' predicted IC50 values of the training set compounds

Compound No	IC50 (nM)	[a]Fit value	Predicted. act	[b]Exp .activity scale	[b]Predicted. activity scale	[c]Reference
1	2	10.12	5	+++	+++	20
64	5	10.1	15	+++	+++	24
73	5	10.14	13	+++	+++	24
67	12	10.02	17	+++	+++	24
71	18	9.79	30	+++	+++	24
69	41	9.96	20	+++	+++	24
77	50	9.86	25	+++	+++	24
80	58	9.48	62	+++	+++	24
56	74	9.22	110	+++	+++	24
57	230	9.19	120	+++	+++	24
81	400	9.61	45	+++	+++	24
20	1700	8.16	1300	++	++	21
58	2000	7.82	2800	++	++	21
10	2000	8.11	1400	++	++	24
3	2300	7.8	2900	++	++	21
9	3800	7.62	4400	++	++	21
22	6500	7.66	4100	+	++	21
6	7600	7.73	3400	+	++	21
15	8700	7.54	5400	+	+	21
44	50000	6.9	23000	+	+	23
47	100000	6.12	140000	+	+	23

[a] Fit value indicates how well the features in the pharmacophore map the chemical features in the compound

[b] Activity scale: +++, IC50 ≤1000 nM (high active); ++, 1000 nM<IC50≤5000 nM (moderate active); +, IC50 >5000 nM (less active)

[c] Molecule data source reference

Table 2 Results of pharmacophore hypothesis generated by using PDF inhibitors

Hypothesis	Total cost	[a]Cost difference	Error cost	RMS	Correlation (r)
1	94.140	59.275	78.695	0.876	0.952
2	94.332	59.083	79.475	0.917	0.946
3	95.220	58.195	80.382	0.963	0.941
4	95.655	57.760	80.586	0.973	0.940
5	95.665	57.750	81.375	1.011	0.934
6	95.892	57.524	82.299	1.053	0.927
7	96.019	57.396	80.983	0.992	0.937
8	83.556	69.859	81.568	1.020	0.933
9	97.472	55.943	82.494	1.062	0.927
10	97.995	55.420	81.980	1.039	0.932

Null cost=153.416; fixed cost=83.5567; configuration cost=11.796

[a] Cost difference=null cost − total cost

Docking studies

The orientation of binding inside the active site of protein may vary if the compounds have diverse chemical space. Moreover, various conformational changes of the active site amino acids may occur during binding of compounds in the pocket. Hence we have used induced fit (IFD; Schrödinger, Inc, USA) [40, 41], a novel method for fast and accurate prediction of ligand induced conformational changes in receptor active sites, for studying the detailed interactions between the ligands and the active site amino acids. The induced fit protocol begins by docking the active ligand with Glide. In order to generate a diverse ensemble of ligand poses, the procedure uses reduced Vander walls radii and an increased Coulomb-vdW cutoff, and can temporarily remove highly flexible side chains during the docking step. For each pose, structure prediction is then used to accommodate the ligand by reorienting nearby side chains.

These residues and the ligand are then minimized. Finally, each ligand is re-docked into its corresponding low energy protein structures and the resulting complexes are ranked according to GlideScore (expanded version of the ChemScore18 scoring function). In our studies, X-ray crystal structure of PDF was taken from PDB entry 1Q1Y and solvent molecules were deleted and bond order for crystal ligand and protein were adjusted and minimized up to 0.30 _ RMSD. Using default settings of IFD which is suitable for a wide range of systems, docking studies was performed on 99 PDF inhibitors.

Results and discussion

The prime objective of the present work is to analyze the structural aspects of PDF and its inhibitors essential higher

Fig. 1 Spatial arrangement of features in 'Hypo 1'
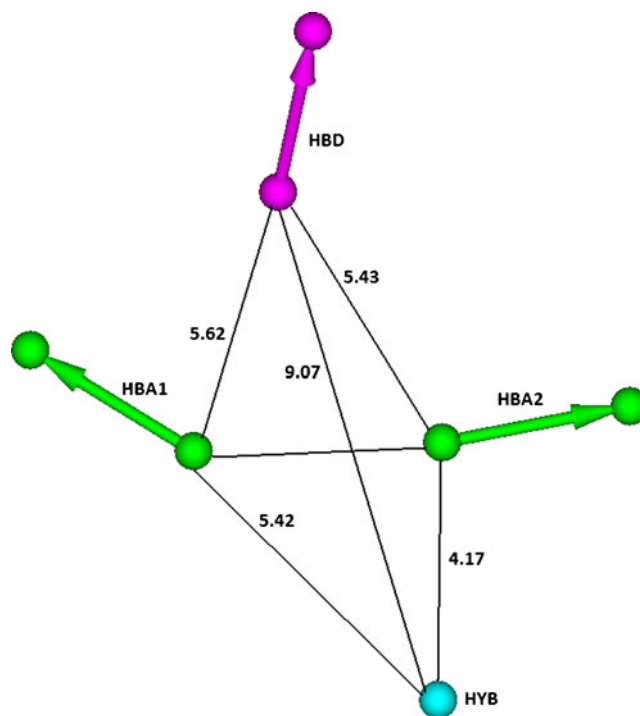
Fig. 2 Distances between the features of 'Hypo 1'

activity and to screen the potential PDF inhibitors from a database. To elucidate the structural features of the inhibitors (pharmacophore), 3D pharmacophore model was generated using a structurally diverse set of compounds with the help of HypoGen module of the Catalyst software. To find the important amino acids (Hot spots) of the deformylase active site involved in ligand binding, induced-fit docking (IFD) [42] docking of Glide software was used.

Pharmacophore model

The chemical space of the training set used in the pharmacophore model generation is illustrated in Chart 1. Using these compounds along with their conformations, ten best quantitative pharmacophore models (hypotheses) were generated (Table 1). The statistical details of these pharmacophores are described in Table 2. Total cost (pharmacophore cost) is calculated for every pharmacophore hypothesis in the top ten generated pharmacophore hypotheses. This cost will signify the inaccuracy and entropy of the individual hypothesis. The model which has less pharmacophore cost is relatively more significant than others. Further it should be very close to the fixed cost (which is an ideal or minimal cost for the given data set). In this study we have obtained a very minimal cost difference, i.e., between fixed cost and the total cost for the 'Hypo 1' (value is 3.8). The null or no correlation cost of the hypotheses signifies the maximum possible error in the activity prediction for the given data set. A value of 40–60 bits difference between null cost and fixed costs for a pharmaco-

Fig. 4 Mapping of features present in 'Hypo 1' with least active compound

phore hypothesis may indicate that it has 75-90% probability of correlating the data. We obtained a cost between the null cost and the total cost value of 59.2758 for 'Hypo 1'. The statistical analysis illustrates that 'Hypo 1' is the best model having more than 90% probability of correlating the data.

As depicted in Table 2, all ten hypotheses showed total cost varying between 94.140 and 97.995 and were very close to the fixed cost (<22 bits). Moreover identical features were obtained for all ten pharamcophores, but their spatial orientation was different. Due to the differences in their special arrangements, they had different statistical values. Among the ten hypothesis generated, 'Hypo 1' had better statistical parameters when compared to others (Table 2). It had low total cost (94.1402), less difference

Fig. 3 Mapping of features present in 'Hypo 1' with high active compound
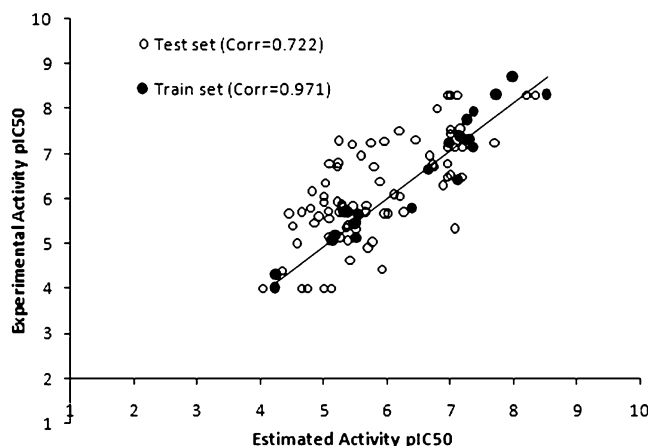
Fig. 5 Scatter plot between the experimental and Hypo1 predicted activity (Test set and Training set correlations are 0.7 and 0.9 respectively)
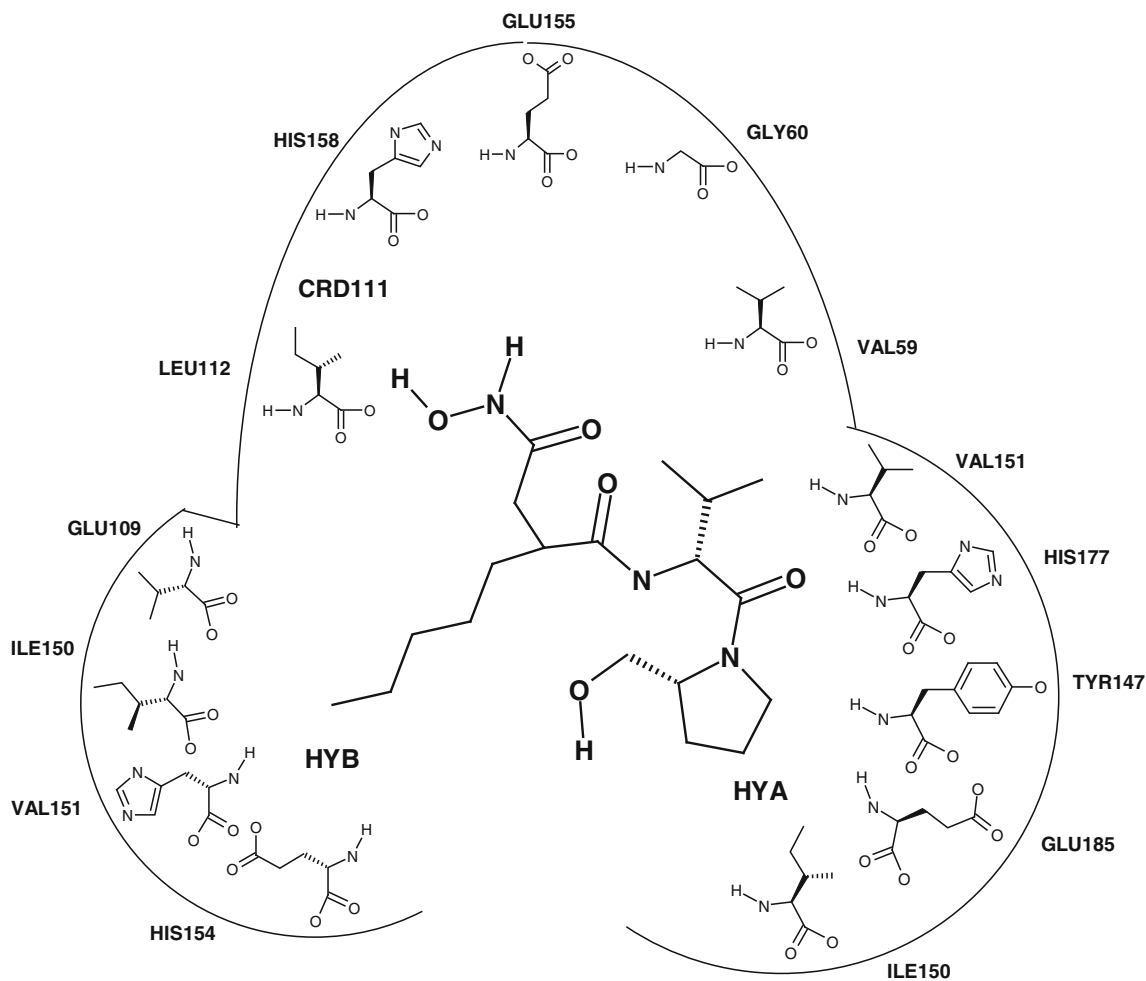
Fig. 6 Schematic diagram of binding modes of high active compounds
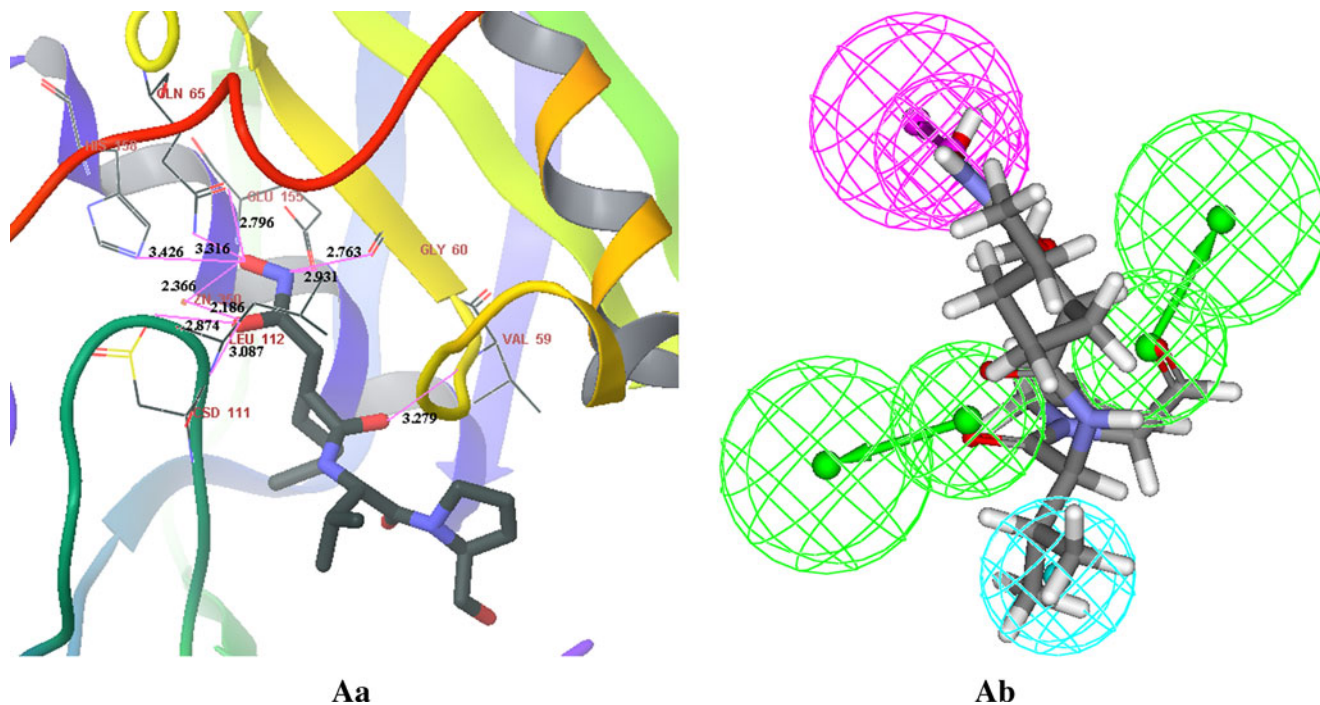
Fig. 7 Binding modes of high active compound

Table 3 Different scaffolds present in total set of molecules, used for induced fit docking
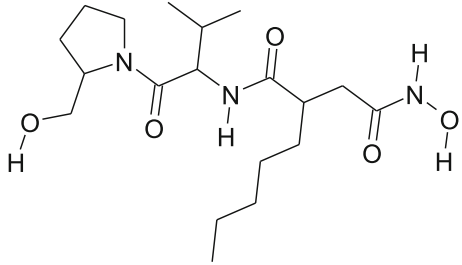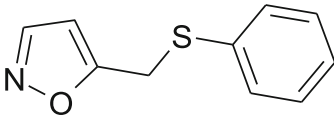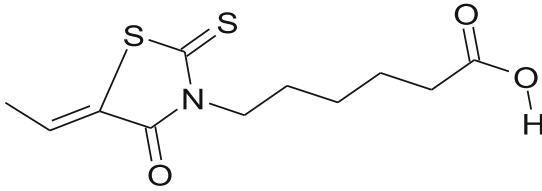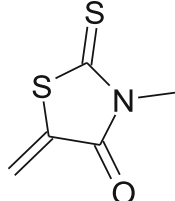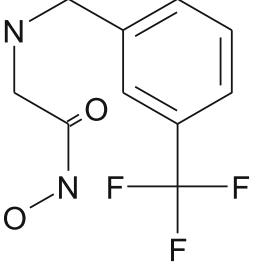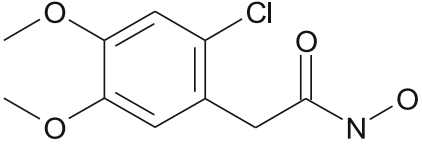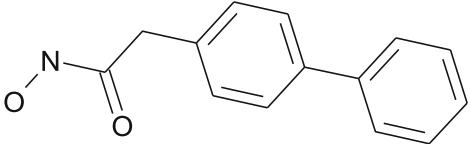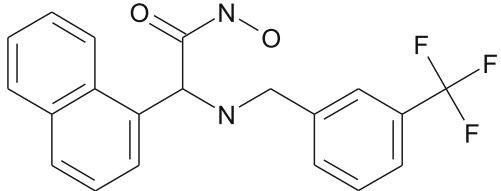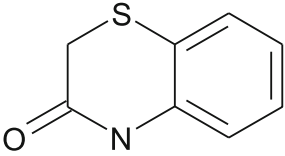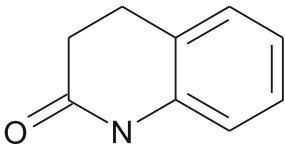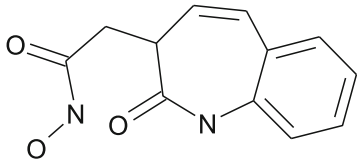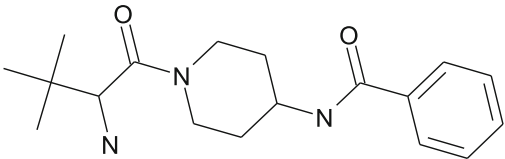
Scaffold	Scaffold ID	No. of mol.	Activity range IC50 (nM)
	A	3	2 - 17
	B	22	800 - 42000
	C	17	890 - 8610
	D	2	10100
	E	3	59 - 690
	F	1	200
	G	1	50000

Table 3 (continued)

Scaffold	Scaffold ID	No. of mol.	Activity range IC50 (nM)
	H	8	46 - 100000
	I	30	5 - 2000
	J	3	63 -175
	K	1	52
	L	11	28 - 200

between total cost and fixed cost (3.8), high cost difference between null cost and total cost (59.276), least RMSD (0.871), and a strong correlation coefficient (0. 952) between experimental and estimated activity. The spatial arrangement of this model consisted of four chemical features, i.e., two hydrogen-bond acceptors (HBA), one hydrogen-bond donor (HBD), and one hydrophobic (HY) feature (Figs. 1 and 2). Activities were estimated for all compounds based on the best ranking pharmacophore (Hypo 1). The actual and estimated PDF activities of the training and test set compounds are listed in Table 1 & Table S1, respectively. As depicted in Table 1, the 'Hypo 1' was able to differentiate the molecules into high, moderate and least actives. All the features of the model are well mapped to the high active compounds and had a fit value ranging from 6.12 to 10.14. Whereas, in the least active

compounds the hydrogen bond donor feature was not mapped and has fit value in the range of 6.1 to 7.5.

The most active compound 1 (Actinonin) had a fitness score of 10.12 when mapped to 'Hypo 1' (Fig. 3). On the contrary, the least active compound 47 was mapped to a value of 6.12 (Fig. 4). In Actinonin, the two HBA features mapped to the two carbonyl oxygen's present in the compound, HBD feature was superimposed to hydroxyl group of CH2OH which is attached to the second position of the pyrrolidenyl ring, which is attached to imidazole ring, and the hydrophobic group is mapped to the alkyl chain of the compound. Whereas, in the least active compound, one of the HBA and a hydrophobic feature were mapped to the compound and the other two features did not map. The predictive power of the 'Hypo 1' was validated with 78 test set compounds. A correlation

coefficient of 0.722, showing a good correlation between the actual and estimated activities (Fig. 5) was obtained. Overall, 32 of 44 highly active, 21 of 28 moderately active, and five inactive compounds were predicted correctly.

Docking studies

Till date, various classes of inhibitors are discovered against *S. aureus* and these were shown in Chart 1 and S1. As PDF/ligand binding interaction models would be useful in designing and optimizing the lead compounds. However, only one experimental evidence is available in the protein databank PDB (1Q1Y) explaining the binding orientation of a particular ligand Actinonin inside the active site of PDF. This scanty and limited information available on the protein-ligand structure complex could explored by docking the known inhibitors into the binding pocket of PDF, to provide understanding of how known PDF inhibitors interact with PDF.

Protein structural flexibility plays a crucial role in receptor–ligand complex formation and ideally should be considered during the drug design process. Moreover the active site of PDF is more flexible due to the presence of large random coil. As the structure-based discovery of new antibacterial agents against *Staphylococcus aureus* would be aided by detailed structural information about the mode of inhibitor binding to its active site, we have used here the flexible docking program of Schrodinger software, induced fit [42] to uncover the conformational variations of PDF active site while interacting with various classes of inhibitors.

The crystal structure of PDF, available in protein databank contains co-crystal actinonin, a potent inhibitor of bacterial PDFs. We have docked all 99 inhibitors (Table S4, Supplementary data) using induced fit docking (IFD) procedure. Briefly, the procedure is intended to

Fig. 9 Alignment of all molecules with their best binding pose

explore the flexibility of both ligand and receptor in the course of the docking process and is composed of three sequential steps: (1) Initial docking of the ligand into the binding pocket of receptor with decreased van der Waals radii of binding site amino acids; (2) optimization of the side chains of the residues in the binding pocket; (3) final docking of the ligand at the optimized receptor. For the docking of 99 inhibitors, the docking region of 1Q1Y was defined as a box with a side of 26 Å and centered on crystal ligand actinonin. Initially the ability of docking method was evaluated by re-docking the crystal ligand actinonin into the binding site of the receptor. Re-docking with actinonin produced best scoring pose with heavy atoms RMSD of the

Fig. 8 Binding pose of potent inhibitor actinonin
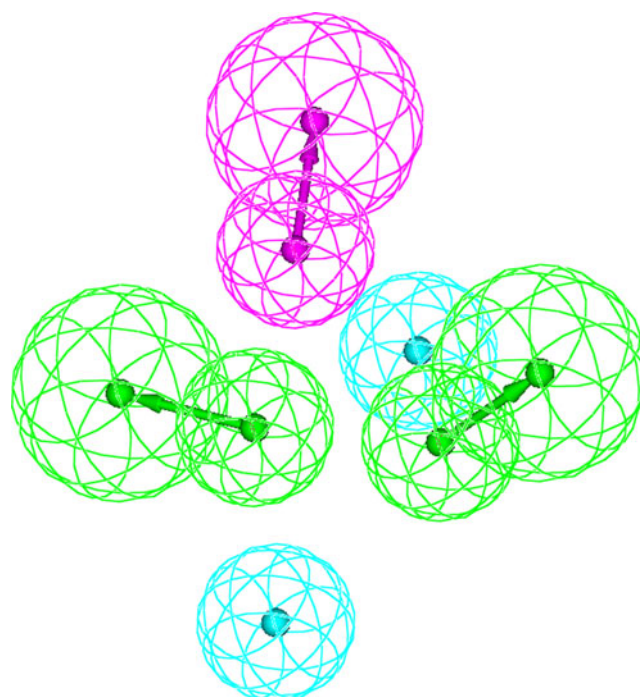
Fig. 10 Modified pharmacophore used for virtual screening

Table 4 Selective hits obtained from virtual screening
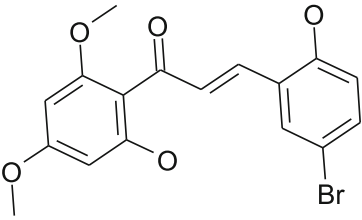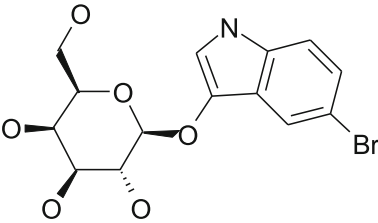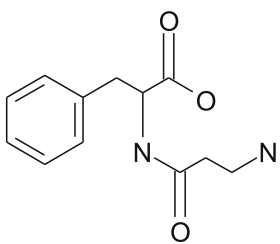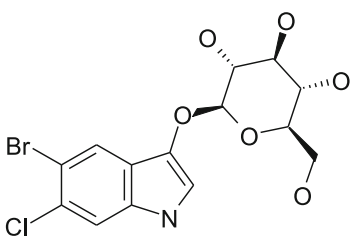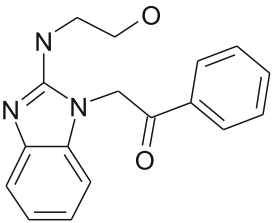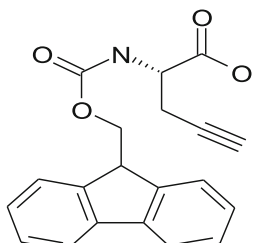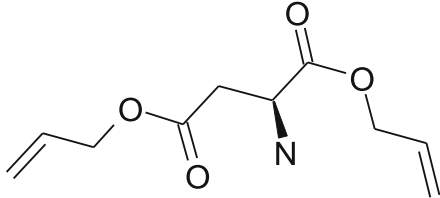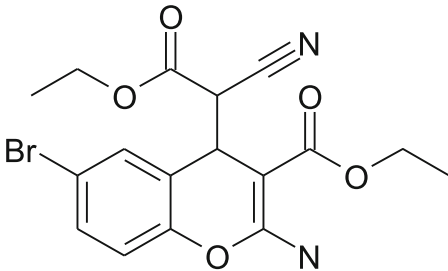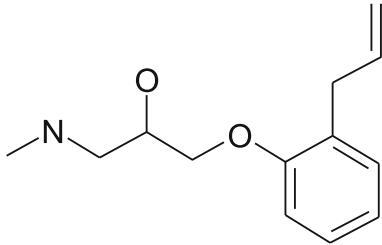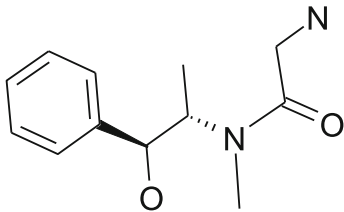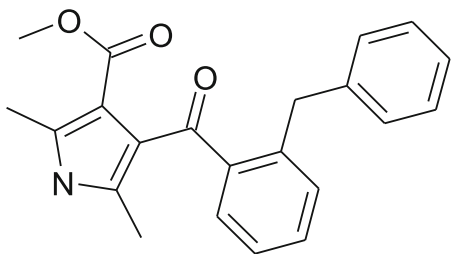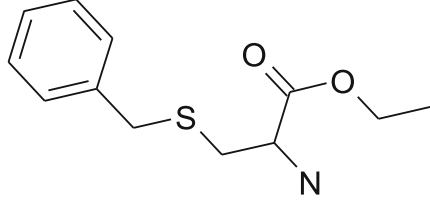
Compound ID	2D-Structure	Predicted activity (nM)	Dock score (CDOCKER Score)
S9720		6.579	45.759
S6052		7.473	50.098
S1245		9.396	51.956
S9473		10.11	46.913
S6231		11.211	40.71
S3880		47.506	46.979

Table 4 (continued)

Compound ID	2D-Structure	Predicted activity (nM)	Dock score (CDOCKER Score)
S2223		70.296	44.986
S9478		185.912	44.066
S3458		193.591	43.738
S2553		216.887	43.682
S6165		328.226	41.12
S2035		402.819	48.414

compound below 2 Å. Figure representing the super positioning of crystal and docked actinonin is available in the Supporting information (Fig S1).

The entire 99 inhibitors were docked into the active site of PDF and correlation was calculated between Glide score and the pIC$_{50}$ by linear regression analysis method. An acceptable correlation coefficient (r) of 0.788 was obtained between experimental pIC$_{50}$ and docking energy (Fig. S2 and Table S2). This correlation proved that the binding conformations and binding models of the inhibitors of PDF are reliable and this is further confirmed by the observation that high active compounds show better glide scores than the low active compounds.

The docking results for all the classes of the compounds are summarized in Table S4 in Supplementary data. The binding modes of high active compounds in each class are shown in Figs. 6, 7, Fig. S2 (Supplementary data) and Table 3. As depicted in Fig. 7a, the amino acids Glu155 and Csd111 (Cystene oxidized to cystene sulfonic acid) create an electronegative environment to form ionic interactions with zinc. All the classes of inhibitors have electronegative groups like hydroxamates or acetates that interact with the positively charged zinc ion. The potent inhibitor actinonin containing electronegative group, hydroxamic acid not only forms a strong ionic interactions with zinc ((O1: 2.36 Å, O2: 2.47 Å) (Fig. 8) but also has strong hydrogen bond interactions with His158 (O1…NE2:3.42 Å), Gln65 (O1… OE1 and NE2: 2.79 Å and 3.32 Å respectively), Csd111 (O2..OD2: 2.874 Å) and with Leu112 (O2…N: 3.087 Å). The NH group of hydroxamic acid also forms strong hydrogen bond interactions with Gln155 (N1….OE2: 2.931 Å) and Gly60 (N1…O:2.763 Å). It also forms another hydrogen bond interaction with its carbonyl oxygen to the back bone NH of Val59 (3.27 Å). The substituted piperdine ring is arranged into another hydrophobic pocket (HYA) and forms van-der Waals interactions with the amino acids, His177, Tyr147, and Glu185. The aliphatic chain, pentyl group of the compound is extended into the hydrophobic region formed by Glu109, Ile150, Val151 and His154 (Figs. 6 and 8).

The compounds which have hydroxamic acid group are showing at least some activity, e.g., compound 45, 48 and 8 by interacting with Zn metal ion, but for higher activity the compounds should have hydrophobic pharmacophores which are missing in these compounds. The hydroxamic group containing compounds like compound 1 and compound 64 show higher activity because their hydrophobic groups are tightly fixed into the hydrophobic pockets A and B (Fig. 7Aa, Ab & Fig. 8) and forming strong hydrophobic interactions with the corresponding amino acids. In the case of compounds 64, 65 and 66, the methyl group present on the nitrogen of the benzo-thiazine ring is showing weak interactions with hydrophobic pocket-A when compared

with iso-pentyl (compound 66) or pentyl (compound 65) functional groups. In the flexible compounds like compound 47 containing carboxyl group (instead of hydorxymate), the orientation of flexible hydrophobic group is different from the other compounds. It extends into another pocket formed by Ser 85, Asp80 and His186, as it cannot be accommodated either in pocket A or B. Due to this extension of the carboxyl group, it is pulled away from the metal ion and has negligible interactions with metal ion. Hence, this may be the reason for having the least activity for these molecules. The alignment of docking poses of the compounds containing different scaffolds were shown in Fig. 7Aa, Ab & Fig. 9. The important features of the compounds necessary for stronger interactions with active site amino acids of PDF are enclosed in circles in this figure.

Like compound 1, the other inhibitors containing hydroxamic acid are also binding in the same orientation. The hydroxamic group compound 41 shows strong electrostatic interactions with metal cation and also with Leu112, Glu156 and Gly60, but its oxazole ring forms weak hydrogen bonds with Val59 and Gly60. Also it is missing important hydrophobic interactions when compared with actinonin. Compound 40 also shows a similar kind of interaction which was observed in 41. Compound 44 contains only one electrophilic group, hydroxamic acid and the rest is bi-cyclic phenyl group. Hence it forms hydrogen bond interactions with Leu112, Glu15 and Gly60. The important interaction with Val59 is missing in this compound. A little change in the orientations of the hydroxamic acid in compound 40, compound 44 and compound 84 is observed. The hydroxamic acid moved slightly far away from Leu112, Glu155 in order to produce a favorable interaction between His154 and oxygen of the methoxy group substituted on phenyl ring of compound 40. The napthyl ring of compound 46 is positioned into the hydrophobic pocket. Hydroxamine and substituted amino
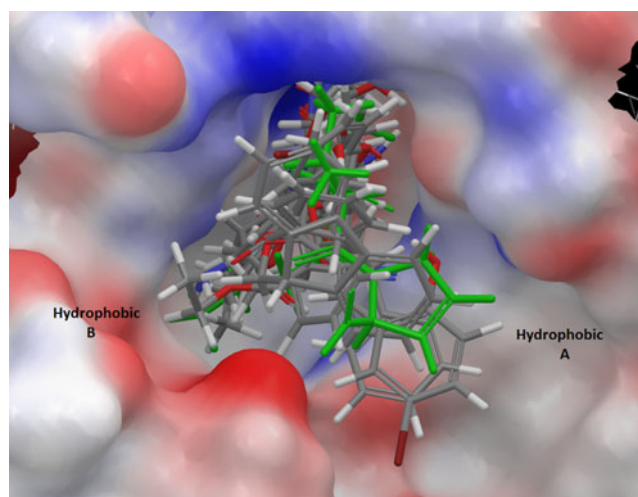
Fig. 11 Binding orientations of screened molecules

phenyl groups are perpendicularly positioned on the napthyl ring to have hydrogen bond interactions with Gly60, Val59 and Cys111. Compound 84 is docked in such a way that it forms an important hydrogen bond between the carbonyl group present on the quinoline ring and Gly50 and Val59. Due to this the hydroxamic position is twisted. Though it is twisted it is conserving the crucial interactions with metal ion and with Leu112. In the non hydroxamic compounds, i.e., in the case of 17 and 39, carboxylic acid is substituted instead of hydroxamic acid. The acidic group forms an electronic interaction with metal cation. In addition, there is hydrogen bonding with Gly60 and Val59 but no observed interaction with the catalytic residue Leu112. They have different orientations of the hydrophobic part when compared with compounds containing hydroxamic acid.

On comparison of the pharmacophore models with docking interaction models, it was found that both are well correlated. The four featured pharmacophore models containing two acceptors, one donor and one hydrophobic feature are mapped well to the best active compound, actinonin. One of acceptors, HBA1 is mapped to carbonyl oxygen of hydroxamine group present in the compound. This group shows important interactions with CSD111, Leu112 and ZN. Another acceptor HBA2 maps to another carbonyl oxygen of the compound, which acts as an important acceptor in forming hydrogen bonds with Val59. The hydrophobic group which is mapped to the iso-pentyl group forms hydrophobic interactions with Glu109, Ile150, Val151 and His154 amino acids. The pharmacophore aligned conformations of diverse compounds are shown in Fig. S3, Supplementary data.

For the inhibitors electro negative group is essential for interacting with the zinc cation ion in the active site of PDF. This is the reason why the simple hydroxamic acid has some activity against PDF. We have collected some of the ligands which interact with Zinc ion from the protein data bank to analyze whether these compounds are able to interact with *S. aureus* PDF (list of compounds along with PDBID was shown in Table S3 in supplementary information). Most of these compounds have -COOH, -CONH2, -CONHOH, -SO2, and -PO4^{-2} which are essential for interacting with zinc metal. These compounds show good predicted activity against *S. aureus* PDF when mapped on to the 'Hypo 1'. All the superimposed compounds on the pharmacophore model were shown in Fig. S4, Supplementary data. These compounds are also having good docking score (Table S3) when these are docked into the active site of PDF. The acidic groups of these compounds are exhibiting strong interactions with zinc metal. The overlay of best poses of these compounds inside the active site of PDF is shown in Fig. S5, Supplementary data. Although hydroxamates are considered the most effective zinc binding group, they lack selectivity, [43] especially they can interact with human physiologically important metallo enzymes. However MMP

inhibitor, doxycycline (Periostat) containing hydroxamate has emerged as an antibiotic in the market. As discussed previously, the PDF inhibitors containing hydroximates can become an antibiotic if they have hydrophobic groups which can properly fit into hydrophobic pockets (HYA and HYB), which are specific for bacterial PDFs.

To select the potent inhibitors with higher selectivity toward PDF, we performed virtual screening. For performing virtual screening of selective inhibitors we have modified the 'Hypo 1' by adding another hydrophobic feature which represents the pocket 'hydrophobic A' of the receptor. The modified model is shown in Fig. 10. Using this pharmacophore model we have screened inhouse database containing 1.5 lacks molecules. A total of 400 compounds were obtained from the pharmacophore screening, among them the compounds having good fitness score (greater than 9.0) and having proper mapping to two hydrophobic features were selected and further docked into the binding pocket of PDF to confirm the conserved interactions for selectivity against PDF. We have selected the best ten molecules which are showing strong hydrogen bonding interactions and also hydrophobic interactions in two hydrophobic pockets of PDF. These compounds were listed in Table 4 and their binding orientations were shown in Fig. 11 *(superimposed all)*.

Conclusions

Ligand based pharmacophore model was generated using 21 diverse PDF inhibitors, best predictive model has two acceptors, one donor and one hydrophobic feature. Further to illustrate the important interactions between the receptor and the inhibitors IndcuedFit docking was performed. Docking results showed that, the electronegative group is essential for the inhibitor which is not only essential for interacting with zinc metal but also with CSD111, Leu112, His158, and Gln65. The two hydrophobic pockets (Hydrophobic A & B) are essential for selectivity toward the PDF. Based on this we have screened inhouse database to find potential and selective inhibitors of PDF. Ten new compounds were obtained from the virtual screenings which have strong interactions with the hydrophobic groups of the receptors.

References

1. Cohen ML (1992) Epidemiology of drug resistance: implications for a post-antimicrobial era. Science 257:1050–1055
2. Hayes JD, Wolf CR (1990) Molecular mechanisms of drug resistance. Biochem J 272:281–290

3. Spratt BG (1994) Resistance to antibiotics mediated by target alterations. Science 264:388–393

4. Adams JM, Capecchi M (1996) N-formylmethionyl-sRNA as the initiator of protein synthesis. Proc Natl Acad Sci USA 55:147–155

5. Adams JM (1968) On the release of the formyl group from nascent protein. J Mol Biol 33:571–589

6. Clements JM, Beckett RP, Brown A, Catlin G, Lobell M, Palan S, Thomas W, Whittaker M, Wood S, Salama S, Baker PJ, Rodgers HF, Barynin V, Rice DW, Hunter MG (2001) Antibiotic activity and characterization of BB-3497, a novel peptide deformylase inhibitor. Antimicrob Agents Chemother 45:563–570

7. Nguyen KT, Hu X, Colton C, Chakrabarti R, Zhu MX, Pei D (2003) Characterization of a human peptide deformylase: implications for antibacterial drug design. Biochemistry 42:9952–9958

8. Apfel C, Banner DW, Bur D, Dietz M, Hirata T, Hubschwerlen C, Locher H, Page MGP, Pirson W, Rosse G, Specklin JL (2000) Hydroxamic acid derivatives as potent peptide deformylase inhibitors and antibacterial agents. J Med Chem 43:2324–2331

9. Apfel C, Banner DW, Bur D, Dietz M, Hubschwerlen C, Locher H, Marlin F, Masciadri R, Pirson W, Stadler H (2001) 2-(2-Oxo-14-dihydro-2 H-quinazolin-3-yl)- and 2-(22-dioxo-14-dihydro-2 H-2lambda6-benzo[1,2,6]thiadiazin-3-yl)-N-hydroxy-acet-amides as potent and selective peptide deformylase inhibitors. J Med Chem 44:1847–1852

10. Thorarensen A, Douglas MR, Rohrer DC, Vosters AF, Yem AW, Marshall VD, Lynn JC, Bohanon MJ, Tomich PK, Zurenko GE, Sweeney MT, Jensen RM, Nielsen JW, Seest EP, Dolak LA (2001) Identification of novel potent hydroxamic acid inhibitors of peptidyl deformylase and the importance of the hydroxamic acid functionality on inhibition. Bioorg Med Chem Lett 11:1355–1358

11. Lee MD, She Y, Soskis MJ, Borella CP, Gardner JR, Hayes PA, Dy BM, Heaney ML, Philips MR, Bornmann WG, Sirotnak FM, Scheinberg DA (2004) Human mitochondrial peptide deformylase a new anticancer target of actinonin-based antibiotics. J Clin Invest 114:1107–1116

12. Xu Y, Lai LT, Gabrilove JL, Scheinberg DA (1998) Antitumor activity of actinonin in vitro and in vivo. Clin Cancer Res 4:171–176

13. Grujic M, Renko M (2002) Aminopeptidase inhibitors bestatin and actinonin inhibit cell proliferation of myeloma cells predominantly by intracellular interactions. Cancer Lett 182:113–119

For examples of hydroxamate and reverse hydroxamate inhibitors see:

14. Apfel C, Banner DW, Bur D, Dietz M, Hirata T, Hubschwerlen C, Locher H, Page MGP (2000) Hydroxamic acid derivatives as potent peptide deformylase inhibitors and antibacterial agents. J Med Chem 43:2324–2331

15. Apfel C, Banner DW, Bur D, Dietz M (2001) 2-(2-Oxo-14-dihydro-2 H-quinazolin-3-yl)- and 2-(22-dioxo-14-dihydro-2 H-2lambda6-benzo[126]thiadiazin-3-yl)-N-hydroxy-acetamides as potent and selective peptide deformylase inhibitors. J Med Chem 44:1847–1852

16. Hackbart CJ, Chen DZ, Lewis JG (2002) N-alkyl urea hydroxamic acids as a new class of peptide deformylase inhibitors with antibacterial activity. Antimicrob Agents Chemother 46:2752–2764

17. Clemens JM, Beckett RP, Brown A (2001) Antibiotic activity and characterization of BB-3497 a novel peptide deformylase inhibitor. Antimicrob Agents Chemother 45:563–570

18. East SP, Beckett RP (2004) Peptide deformylase inhibitors with activity against respiratory tract pathogens. Bioorg Med Chem Lett 14:59–62

19. Hu X, Nguyen KT (2003) Structure-based design of a macrocyclic inhibitor for peptide deformylase. J Med Chem 46:3771–3774

20. Serero A, Giglione C, Sardini A, Martinez Sanz J, Meinnel T (2003) An unusual peptide deformylase features in the human mitochondrial N-terminal methionine excision pathway. J Biol Chem 278:52953–52963

21. Yuan Z, Trias J, White RJ (2001) Deformylase as a novel antibacterial target. Drug Discov Today 6:954–961

22. Jain R, Chen D, White RJ, Patel DV, Yuan Z (2005) Bacterial peptide deformylase inhibitors: a new class of antibacterial agents. Curr Med Chem 12:1607–1621

23. Lee MD, Antczak C, Li Y, Sirotnak FM, Bornmann WG, Scheinberg DA (2003) A new human peptide deformylase inhibitable by actinonin. Biochem Biophys Res Commun 312:309–315

24. Giglione C, Pierre M, Meinnel T (2000) Peptide deformylase as a target for new generation, broad spectrum antimicrobial agents. Mol Microbiol 36:1197–1205

25. Boularot A, Giglione C, Petit S, Duroc Y, Alves de Sousa R, Larue V, Cresteil T, Dardel F, Artaud I, Meinnel T (2007) Discovery and refinement of a new structural class of potent peptide deformylase inhibitors. J Med Chem 50:10–20

26. Cali P, Naerum L, Mukhija S, Hjelmencrantz A (2004) Isoxazole-3-hydroxamic acid derivatives as peptide deformylase inhibitors and potential antibacterial agents. Bioorg Med Chem Lett 14:5997–6000

27. Michael HH, Cenizal T, Gutteridge S, Wayne SH, Tao Y, Totrov M, Vernon AW, Zheng Ya-jun (2004) A novel class of inhibitors of peptide deformylase discovered through high-throughput screening and virtual ligand screening. J Med Chem 47:6669–6672

28. Atli T, Martin RD, Douglas CR, Anne FV (2001) Identification of novel potent hydroxamic acid inhibitors of peptidyl deformylase and the importance of the hydroxamic acid functionality on inhibition. Bioorg Med Chem Lett 11:1355–1358

29. Valentina M, Xiaohui H, Juliet N, Kunyong Y, Andreas K (2004) Identification of novel potent bicyclic peptide deformylase inhibitors. Bioorg Med Chem Lett 14:1477–1481

30. Seung WY, Hee YL, Bong HC, Kyung MA, Jung SR, Young HL, Jae HK (2006) Synthesis and biological evaluation of N-Formyl hydroxylamine derivatives as potent peptide deformylase inhibitors. Bull Korean Chem Soc 27:1075–1078

31. Jung SY, Chang JZ, Sangku L, Jin HK, Won GK (2006) Macrolactin N, a new peptide deformylase inhibitor produced by Bacillus subtilis. Bioorg Med Chem Lett 16:4889–4892

32. CATALYST 4.10, Accelrys, Inc, San Diego, CA, 2005, http://www.accelrys.com/

33. Kurogi Y, Guner OF (2001) Pharmacophore modeling and three-dimensional database. Curr Med Chem 8:1035–1055

34. Funk OF, Kettmann V, Drimal J, Langer T (2004) Chemical function based pharmacophore generation of endothelin-A selective receptor antagonists. J Med Chem 47:2750–2760

35. Brooks BR, Brucolleri RE, Olafson BD, States DJ, Swaminathan S, Karplus MJ (1983) CHARMM: a program for macromolecular energy. J Comput Chem 4:187–217

36. Guner OF (2000) Pharmacophore perception development and use in drug design. International University Line, La Jolla

37. Sprague PW (1995) In: Muller K (ed) Perspectives in drug discovery and design, vol 3. ESCOM Science Publishers BV, The Netherlands, p 1

38. Smellie A, Teig SL, Towbin P (1995) Poling-promoting conformational variation. J Comput Chem 16:171–187

39. Santo RD, Fermeglia M, Ferrone M, Paneni MS, Costi R, Artico M, Roux A, Gabriele M, Tardif KD, Siddiqui A, Pricl S (2005) Simple but highly effective three-dimensional chemical-

feature-based pharmacophore model for diketo acid derivatives as hepatitis C virus RNA-dependent RNA polymerase inhibitors. J Med Chem 48:6304–6314

40. 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

41. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem 47:1750–1759

42. Sherman W, Day T (2006) Novel procedure for modelling ligand/receptor induced fit effects. J Med Chem 49:534–553

43. Docherty AJ, Crabbe T, O'Connell JP, Groom CR (2003) Proteases as drug targets. Biochem Soc Symp 70:147–161

ORIGINAL PAPER

CYP isoform specificity toward drug metabolism: analysis using common feature hypothesis

M. Ramesh · Prasad V. Bharatam

Abstract Three dimensional pharmacophoric maps were generated for each isoforms of CYP2C9, CYP2D6 and CYP3A4 separately using independent training sets consist of highly potent substrates (seven substrates for each isoform). HipHop module of CATALYST software was used in the generation of pharmacophore models. The best pharmacophore model was chosen out of the several models on the basis of (i) highest ranking score, (ii) better fit value among training set, (iii) capability to screen substrates from data set and (iv) efficiency to identify the isoform specificity. The individual pharmacophore models (CYP2C9-hypo1, CYP2D6-hypo1 and CYP3A4-hypo1) are characterized by the pharmacophoric features XZDH, RPZH and XYZHH for the CYP2C9, CYP2D6 and CYP3A4 respectively. Each of the chosen models was validated by using data sets of CYP substrates. This comparative study of CYP substrates demonstrates the importance of acidic character along with HBD and HBAl features for CYP2C9, basic character with ring aromatic features for CYP2D6 and hydrophobic features for CYP3A4. Acidity, basicity and hydrophobicity features arising from the functional groups of the substrates are also responsible for demonstrating CYP isoform specificity.

M. Ramesh · P. V. Bharatam (✉)
Department of Medicinal Chemistry, National Institute of
Pharmaceutical Education and Research (NIPER),
Sector-67, S. A. S Nagar,
Mohali 160 062, India
e-mail: pvbharatam@niper.ac.in

P. V. Bharatam
e-mail: bharatampv@yahoo.com
URL: http://pvbharatam.net/

Hence, these chemical features are incorporated in the decision tree along with pharmacophore maps. Finally, a decision tree based on chemical features and pharmacophore features was generated to identify the isoform specificity of novel query molecule toward the three isoforms.

Introduction

Drugs, in many cases, are metabolized into more readily soluble, excretable forms and are eliminated from the systemic circulation. The role of drug metabolism is detoxification and potential excretion of harmful compounds to nontoxic compounds, though in a few cases metabolism transforms nontoxic compounds to toxic compounds. All of these reactions are significantly mediated by cytochromes P450 commonly known as CYPs [1, 2]. The CYPs constitute a super family of heme-containing monooxygenase enzymes that catalyze the metabolism of a wide variety of endogenous and exogenous compounds. Predominant isoforms of cytochrome P450 involved in drug metabolism are CYP2C9, CYP2D6 and CYP3A4 which metabolize up to 80-90% of all xenobiotics [3–6]. CYP2D6 metabolizes 15-20% of clinically administered drugs. But the response arises due to this isoenzyme varies from patient to patient because of its associated genetic polymorphism. CYP3A4 metabolizes up to 50% of all reported drugs and a few of which are also metabolized by other isoforms. Though there are commonalities in the oxidation reactions carried out by various isoforms of cytochromes, isoform specificity is also observed. Isoform specificity of

cytochromes is manifested in any of the following ways- (i) substrate selection- i.e., specific substrate predominantly metabolized by specific enzyme, it may be due to specific interactions between the substrate and enzyme (e.g., flurbiprofen metabolized by CYP2C9 through ionic interactions) [7] (ii) Site of metabolism- i.e., specific substrate may be oxidized by more than one isoenzyme but at two different locations on the substrate, (e.g., 4' hydroxylation is the site of metabolism (SOM) of diclofenac in CYP2C9 whereas 5-hydroxylation is the site of metabolism in CYP3A4) [8] (iii) Rate of conversion to product- i.e., while a specific substrate may be oxidized by two different isoenzymes and when the site of metabolism is the same, the K_m value associated with catalytic process differs, (e.g., fluoxetine shows higher K_m value (K_m=30.7 μM) for CYP2C9 and lower K_m value (K_m=2.20 μM) for CYP2D6) [9].

CYP2D6 shows genetic polymorphism and pharmaceutical industries are less inclined to develop a molecule with CYP2D6 mediated drug metabolism. Binding of any drug to CYP2D6 allelic variants cause drug-drug interactions [10]. The metabolic profile of ligand molecule also varies from one cytochrome to another [8]. It is important to establish which isoforms play a role in metabolizing any new drug candidate. Hence, the prediction of isoform specificity is required at the early stage of drug discovery to optimize the metabolic profile of new leads. Prediction of isoform specificity also helps to switch the ligand metabolism toward specific CYPs in order to avoid the formation of unwanted/toxic metabolites from nonselective CYPs. There are several experimental drug metabolism studies to perform the isoform specificity using LC/MS. All these experimental drug metabolism studies are time consuming and expensive. There is a need to develop strategies for identifying the isoform specificity in metabolizing a new drug candidate in order to reduce the number of experimental drug metabolism studies. Computational metabolite prediction studies offer such strategies for the prediction of isoform specificity and metabolites on drug metabolism. Early identification of isoform specificity and metabolites on drug metabolism through the computational strategies can also help in the design of novel chemical scaffolds with efficient metabolic profile. A few scientific groups proposed descriptor based classification schemes for identifying CYP isoform specificity [11–14]. The aim of the current work is to develop a schematic protocol based on pharmacophore maps which can identify the preferred enzymes for the metabolism of a given substrates among CYP2C9, CYP2D6 and CYP3A4. This information based on computational studies coupled with knowledge on drug metabolites could be a valuable tool in the design analogues with improved metabolic profile within a specific class of therapeutic agents.

Various reports on the computational pharmacophore models for CYP2C9, CYP2D6 and CYP3A4 are briefly reviewed below. De Groot et al. reported a combination of protein and pharmacophore modeling from 31 metabolic pathways of 27 substrates for CYP2C9. The model also used molecular orbital calculations on substrates that incorporated steric, electronic and chemical stability properties. The model was successfully used to predict the metabolism of substrates [15]. Jones et al. built the initial model for CYP2C9 and the model was further updated by overlaying eight substrates and one inhibitor, using phenytoin as the template molecule [16]. Mancy et al. constructed a model which demonstrated a hydrogen bond donor–hydrogen bond acceptor interaction and resulted in a relatively larger active site cavity for CYP2C9 [17]. The substrate models for CYP2D6 were constructed by manual alignments based on substrates containing a basic nitrogen atom at either 5Å or 7Å from the site of oxidation, and coplanar aromatic rings [18, 19]. Another hypothetical model developed by Koymans et al. used debrisoquine and dextromethorphan as template molecules. The model constructed from 16 substrates representing 23 metabolic reactions and was validated by predicting the metabolism of a further four compounds [20]. Lewis et al. suggested the structural requirements for CYP3A4 substrates which include hydrogen bond acceptor atom 5.5–7.8Å from the site of metabolism [21]. Ekins et al. generated a hypothesis for CYP3A4 from 38 substrates consisting of two hydrogen bond acceptors, one hydrogen bond donor and one hydrophobic region [22]. There are various hypothetical models and computational drug metabolism studies [23–25] which independently help in understanding metabolic profile of CYPs. However, a combined protocol has not yet been established to identify the isoform specificity of CYP substrates. In this article, we describe a combined protocol using pharmacophore maps and chemical features.

Materials and methods

All the computational studies using softwares SYBYL6.9 [26] and CATALYST4.10 [27] were carried out on Silicon Graphics Octane2 workstation, running under the IRIX 6.5 operating system.

HipHop is an automated tool within CATALYST software that is based on alignment of common features present in highly potent compounds, it is a useful computational tool in building three-dimensional pharmacophore based hypothetical models from the highly potent molecules of diverse conformational structures. It employs a polling algorithm to identify the common features among the training set of compounds and this approach does not

Table 1 Training set substrates for HipHop hypotheses generation
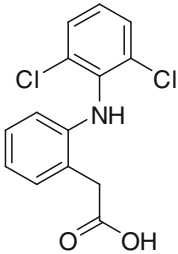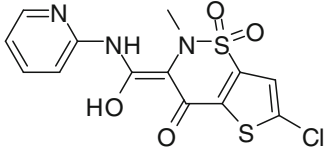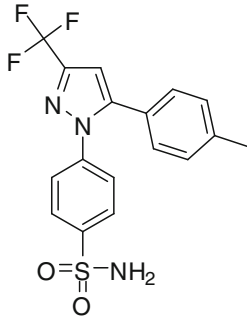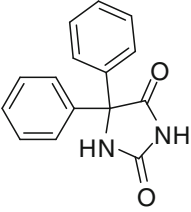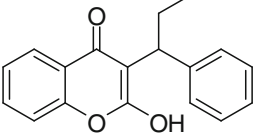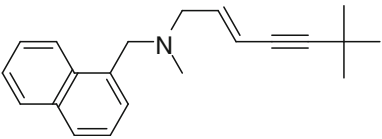
Training set substrates	Structures	K_m Values (µM)	Fit value	Ref.
	CYP2C9			
Diclofenac (**1**)		1.4	3.05	[37]
Lornoxicam (**2**)		0.83	2.98	[38]
Celecoxib (**3**)		4.10	1.93	[39]
Phenytoin (**4**)		5.40	2.69	[40]
Phenprocoumon (**5**)		0.72	4.00	[41]
Terbinafine (**6**)		3.20	2.84	[42]

Table 1 (continued)

Training set substrates	Structures	K_m Values (µM)	Fit value	Ref.
Warfarin (7)		1.5	3.93	[43]

CYP2D6

7-methoxy-4-[(methylamino)methyl]-2H-chromen-2-one (8)		5.09	3.99	[44]
Citalopram (9)		7.0	2.62	[45]
Astemizole (10)		0.96	2.99	[46]
Venlafaxine (11)		10.0	3.73	[47]
Loperamide(12)		2.8	3.54	[48]
4-((ethylamino)methyl)-7-methoxy-2H-chromen-2-one (13)		1.00	4.00	[44]
Amitriptyline (14)		5.4	3.72	[49]

Table 1 (continued)
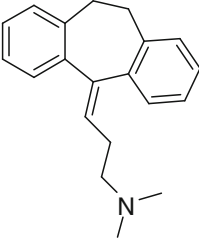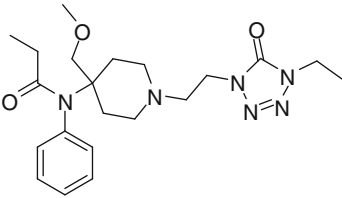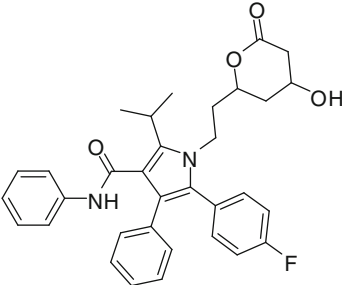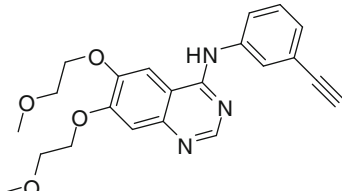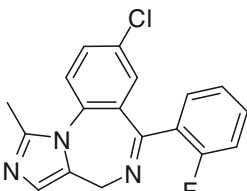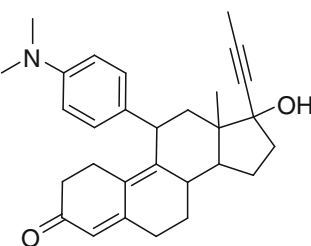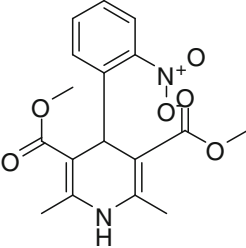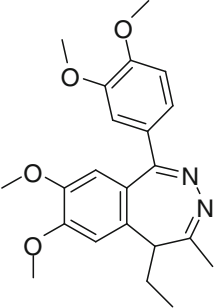
Training set substrates	Structures	K_m Values (µM)	Fit value	Ref.
	CYP3A4			
Alfentanil (**15**)		1.0	3.87	[50]
Atorvastatin lactone (**16**)		1.4	3.83	[51]
Erlotinib (**17**)		5.92	3.62	[52]
Midazolam (**18**)		8.44	2.90	[53]
Mifepristone (**19**)		7.4	3.54	[54]

Table 1 (continued)

Training set substrates	Structures	K_m Values (µM)	Fit value	Ref.
Nifedipine (**20**)		3.2	3.78	[55]
Tofisopam (**21**)		0.9	4.91	[56]

consider the biological activity while generating hypotheses. The usefulness of this pharmacophore modeling approach has already been established by numerous successful applications of our studies [28, 29].

Collection of CYP2C9, CYP2D6 and CYP3A4 substrates

In this study, we have collected the data of known CYP2C9, CYP2D6 and CYP3A4 substrates from publicly available literature. The substrates were collected on the basis of availability of experimental information on CYP catalyzed product formation. CYP2C9, CYP2D6 and CYP3A4 substrates which meet the following conditions in the primary literature are included in three different data sets- (i) Kinetic data (K_m /V_{max}) exists for metabolism by recombinant CYP2C9, CYP2D6 and CYP3A4 respectively, (ii) Sterio-chemical details are defined, (iii) Structure of the metabolite is clearly determined and (iv) The site of metabolism is established using experimental metabolism studies. A few substrates (fluoxetine, troglitazone, fluvastatin, celecoxib, tolterodine, thioridazine, astemizole, delavirdine, tamoxifen) are part of more than one data set as they are reported to be catalyzed by more than one isoforms. These three data sets contain 30, 37 and 74 numbers of substrates respectively for CYP2C9, CYP2D6 and CYP3A4.

Training set and test set selection

The substrates contained in the training set were selected by considering their therapeutic diversity as well as their

structural diversity in the form of core and side chain of the molecule. The seven most potent substrates were selected for each of the isoforms (Table 1). The data sets defined in the previous section are employed as test sets, these include substrates chosen for training set.

Ligand pre-processing and conformational model generation

The 3D structure of CYP substrates were built using *sketch molecule* panel implemented in SYBYL6.9 and were energetically minimized by Powell method. The optimized 3D structures were stored as .mol2 files and transferred to CATALYST workstation for the conformational model generation. Maximum of 250 conformations were allowed to generate a conformational model with an energy cut of value 20 kcal mol^{-1}. The number of conformations needed to produce a good representation of a compound's conformational space depends on the structure of the molecule. The conformational diversity of the training set substrates also accounts for the molecular flexibility of the ligand while binding to the active site of the enzymes.

Functional mapping analysis

The 3D structures of the substrates were analyzed for the functional mapping individually to ascertain the presence of different pharmacophoric features in the training set substrates. CATALYST software provides two types of opportunities to carry out functional

Table 2 Top ranked hypotheses

CYP	Hypotheses rank	Features	Scores
2C9	hypo1	XZDH	48.02
	hypo2	XZDH	47.34
	hypo3	RZHH	46.63
	hypo4	RZHA	45.63
	hypo5	RZHA	45.47
	hypo6	XXDH	44.92
	hypo7	RXHH	44.73
	hypo8	XZHH	44.37
	hypo9	RXHA	43.73
	hypo10	XXDA	43.72
2D6	hypo1	RPZH	59.93
	hypo2	RPZH	59.93
	hypo3	RPZA	58.73
	hypo4	PXZH	58.73
	hypo5	RPZA	58.67
	hypo6	RPYH	57.58
	hypo7	RPYH	57.58
	hypo8	RPYH	57.32
	hypo9	PXYH	56.40
	hypo10	RPYA	56.37
3A4	hypo1	XYZHH	88.89
	hypo2	XYZHA	87.49
	hypo3	YZZHH	83.29
	hypo4	XZZHH	83.29
	hypo5	YZZZHH	83.10
	hypo6	XYZHH	83.01
	hypo7	XYZHA	82.93
	hypo8	XYZHH	82.92
	hypo9	XYZHA	82.73
	hypo10	RYZHH	82.59

Summary of the features definitions

A: HBA; H: HBA lipid; D: HBD; Z: HYDROPHOBIC; Y: HYDROPHOB aliphatic; P: Posionizable

X: HYDROPHOB aromatic; R: RING AROMATIC

mapping analysis in the form of (i) fragment based functional mapping (ii) features based functional mapping. Feature based functional mapping analysis was performed for all the training set substrates. This helped us to incorporate the required features from the feature dictionary while generating the common feature hypotheses. Pharmacophoric features like hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), hydrogen bond acceptor lipid, hydrophobic, hydrophobic aliphatic, hydrophobic aromatic and ring aromatic were searched among the training set substrates using feature functional mapping.

HipHop hypotheses generation

HipHop provides feature based alignment of collection of compounds without considering biological activity. It matches pharmacophoric features of the molecules against template. HipHop takes collection of conformational models of molecules and produces series of molecular alignments in variety of standard file formats. HipHop begins by identifying configurations of features common to a set of molecules. A configuration consists of a set of relative locations in 3D space and associated feature types. A molecule matches the configuration if it possesses the conformations and pharmacophoric features that can be superimposed within a tolerance limit from the corresponding ideal locations.

The training set compounds associated with their diverse conformational models were subjected for hypotheses generation. The molecular characteristics, which are essential for tight binding between ligand and its corresponding targets like CYP2C9, CYP2D6 and CYP3A4 were expressed as common features disposed in three-dimensional space known as hypotheses. Phenprocoumon is the highly potent CYP2C9 substrate in the training set used as the principal compound for hypotheses generation. Phenprocoumon, 7-methoxy-4-[(methylamino)methyl]-2*H*-
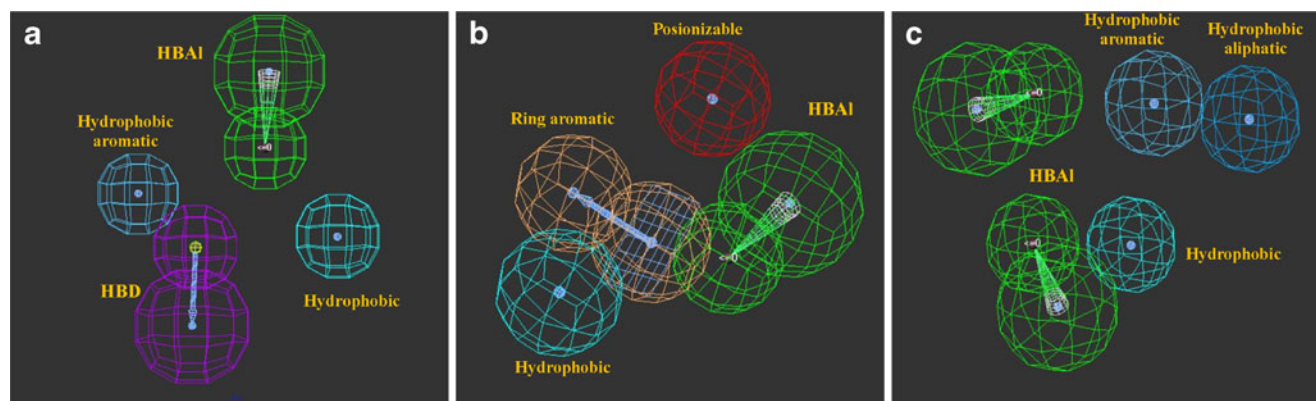
Fig. 1 Top ranked common features hypotheses (HipHop) for **a**) CYP2C9 **b**) CYP2D6 **c**) CYP3A4

chromen-2-one and tofisopam were defined as principal compounds respectively for CYP2C9, CYP2D6 and CYP3A4 during the spreadsheet generation. While generating hypotheses customization to feature requirement in model generation are possible according to each situation, so that not all the substrates have to match all the features in the proposed hypothesis. The complete misses, feature misses and misses option were optimized during hypotheses generation. The pharmacophoric surface accessible features like hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), hydrogen bond acceptor lipid, hydrophobic, hydrophobic aliphatic, hydrophobic aromatic and ring aromatic etc. were included from the feature dictionary while generating HipHop hypotheses. The jobs were submitted separately for individual isoforms to generate hypotheses in CATALYST software and carefully monitored until the valid maps are generated.

Results and discussion

Overview of CYP2C9, CYP2D6 and CYP3A4 HipHop hypotheses

CATALYST software returned 10 possible pharmacophore hypotheses (Table 2) for each isoform and these solutions have different arrangement of constituent features and ranking scores. Each and every hypothesis was manually visualized and critically analyzed. After deleting the redundant features of hypotheses that have the same chemical characteristics and nearly the same distance between features, diverse configurations of hypotheses were selected. All the CYP2C9 hypotheses contain four features with the ranking scores ranging from 48.02 to 43.72. Hypotheses-1 and 2 consist of the same common feature functions of hydrophobic aromatic, hydrophobic, hydrogen bond donor (HBD) and hydrogen bond acceptor lipid (HBAl). CYP2D6 hypotheses also contain four features with the ranking scores ranging from 59.93 to 56.37. Top two ranked hypotheses consist of the same common feature functions of ring aromatic, posionizable, hydrophobic and HBAl features.

The top four ranking hypotheses of CYP3A4 contain five features. CYP3A4 hypotheses ranking scores vary from 88.89 to 82.59. The top ranked hypothesis of CYP3A4 consists of hydrophobic aromatic, hydrophobic aliphatic, hydrophobic and HBAl as the common-feature functions.

Mapping of training set substrates on the top ranked HipHop hypothesis

The individual training set compounds were fitted on to the top ranked hypothesis (Fig. 1). The fitting was carried out using compare fit option implemented in CATALYST software. Best fit option was used for obtaining the fit value (Table 1) which also provides information regarding the best possible conformer of the substrate. The most active CYP2C9 substrate phenprocoumon (5) mapped on to all the features of the HipHop hypothesis. Structural fragments like hydroxyl (OH), carbonyl (C=O), coumarin ring and phenyl ring of (5) mapped on to HBD, HBAl, hydrophobic and hydrophobic aromatic features respectively. Important features such as HBD and HBAl in CYP2C9-hypo1 of the surface-accessible models were found to be optimal for CYP2C9 substrate specificity. They are the minimum components of hypothesis for effective CYP2C9 binding affinity. Another CYP2C9 substrate warfarin (7) also contributes with all four pharmacophoric features with a fit value of (3.93). Highest fit value (4.00) was observed for 4-((ethylamino) methyl)-7-methoxy-2 H-chromen-2-one (13) and lowest fit value (2.62) was observed for citalopram (9) on CYP2D6-hypo1 (Fig. 2). Highest fit value on CYP3A4-hypo1 was shown by the most active compound tofisopam (21) among CYP3A4 training set. As the number of overlapping features reduces, the fit value also observed to decrease indicating the robustness of the chosen model.

Validation of three top ranked HipHop pharmacophore models

A satisfactory pharmacophore model not only should be able to predict the pharmacophoric features of the

Fig. 2 Superimposition of (**a**) 4-((ethylamino) methyl)-7-methoxy-2H-chromen-2-one (highest fit value) (**b**) Citalopram (lowest fit value) on the validated pharmacophore map (CYP2D6-hypo1) of CYP2D6
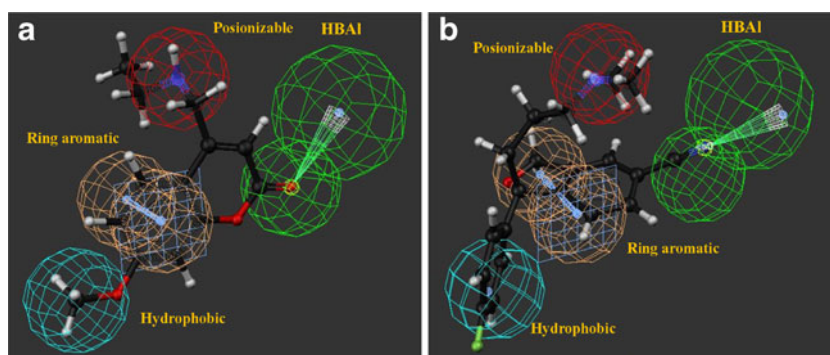
Table 3 Predicting isoform specificity using chemical features

No. of substrates in data sets	Isoform specificity: Chemical features		
	CYP2C9: Group-1	CYP2D6: Group-2	CYP3A4: Group-3
CYP2C9 (total: 30)	27/30	3/30	12/30
CYP2D6 (total: 37)	5/37	34/37	17/37
CYP3A4 (total: 74)	29/74	23/74	63/74

training set compounds correctly, but also must be capable of predicting the pharmacophoric features of external compounds present in the test set. Three independent data sets comprised of 30, 37 and 74 substrates including the molecules involved in model building was used to validate the established model of CYP2C9-Hypo1, CYP2D6-Hypo1 and CYP3A4-Hypo1 respectively. The top ranked CYP2C9-Hypo1, CYP2D6-Hypo1 and CYP3A4-Hypo1 respectively screened 23, 29 and 71 substrates from their corresponding data set. The success rate in screening test set of substrates has been found to be above 75%, thus the top ranked model of CYP2C9, CYP2D6 and CYP3A4 are considered as valid models. This high success rate establishes the reliability of the pharmacophore models. The pharmacophore models (CYP2C9-hypo1, CYP2D6-hypo1 and CYP3A4-hypo1) were finalized as best hypotheses for CYP2C9, CYP2D6 and CYP3A4 respectively by considering their highest ranking scores, better fit value among their training set and successful screening of substrates from their corresponding data sets.

Role of common chemical features and pharmacophoric features among CYP specific substrates

Pharmacophoric maps developed in the previous section were based on the generalized pharmacophoric features like hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), hydrophobic, hydrophobic aliphatic, hydrophobic aromatic and ring aromatic etc. Generalized features do not incorporate the chemical features like weak acidity neither in the form of pharmacophoric features nor in the form of physicochemical parameterization. Hence, the pharmacophoric map alone cannot be exclusively useful in identifying the isoform specificity since access channel and active site of CYPs are controlled by particular acidic, basic or hydrophobic residues. Acidic, basic and hydrophobic chemical features of

CYP substrates can be a minimum requirement for the initial entry of the substrates through the access channel.

Access channel and active site of CYP2C9 consists basic residues (Lys72, Arg97, Arg108 and Asn204), this reveals the importance of weak acidic character (Group-1) among CYP2C9 substrates. Substitution of other amino acids for Arg108 has been shown to diminish greatly the oxidation of diclofenac [30–32]. Access channel (Glu216) and active site (Asp301) of CYP2D6 constitutes acidic residues which reveal the importance of basic character among CYP2D6 substrates [10]. An independent investigation also suggests that Asp301 plays a structural role [33] and the result indicates that specificity for the acidic residues Asp301 and Glu216 requires basic nitrogen [34]. The proposed key role and importance of Glu216 was also confirmed by experiment [35]. The mutation of Glu216 also altered the specificity. This suggests that the binding site of CYP2D6 favors the basic substrates. This was also verified when the crystal structure of CYP2D6 was determined [10]. Similarly, the access channel and active site of CYP3A4 constitutes the hydrophobic residues (Phe) [36], reveals the importance of hydrophobic character observed in CYP3A4-hypo1. Acidic character of CYP2C9 substrates, basic character of CYP2D6 substrates and hydrophobic character of CYP3A4 substrates need to be considered while predicting the isoform specificity which can be a minimum requirement for the substrates to be CYP specific. The substrates having necessary pharmacophoric features may fail to reach the active site, if they do not have sufficient acidic/basic/hydrophobic feature and similarly the substrates having ease of access may fail to bind at the active site if they do not possess necessary pharmacophoric features.

Chemical features which determine the accessibility to particular CYPs in combination with pharmacophoric features which determine the binding affinity at the active site of particular CYPs may better help in identifying the isoform specificity of CYP substrates. Hence, it was found necessary

Table 4 Predicting isoform specificity using pharmacophore maps

No. of substrates in data sets	Isoform specificity: Hypotheses		
	CYP2C9-hypo1	CYP2D6-hypo1	CYP3A4-hypo1
CYP2C9 (total: 30)	23/30	4/30	19/30
CYP2D6 (total: 37)	13/37	29/37	15/37
CYP3A4 (total: 74)	45/74	33/74	71/74

Table 5 Validation of decision tree to identify the isoform specificity

No. of substrates in data sets	Isoform specificity: Decision tree		
	CYP2C9-hypo1	CYP2D6-hypo1	CYP3A4-hypo1
CYP2C9 (total: 30)	23/30	2/30	7/30
CYP2D6 (total: 37)	3/37	29/37	6/37
CYP3A4 (total: 74)	20/74	22/74	61/74

to consider the combination of chemical features as well as pharmacophoric features in developing a protocol to identify the isoform specificity. To account, for the influence of chemical features, substrates considered in this work were segregated into three groups on the basis of acidic, basic and hydrophobic chemical features. These chemical features which are grouped into group-1, group-2 and group-3 preferentially are suitable to identify the isoform specificity of CYP2C9, CYP2D6 and CYP3A4 respectively. The presence of one of these chemical features, gives a clue to identify the isoform specificity by carrying out screening at particular hypothesis rather than screening by all three hypotheses. Presence of these chemical features in the ligand also gives information regarding the possibilities of cytochromes to be involved in drug metabolism. This information will be useful at the early stage of drug discovery to switch the cytochrome mediated drug metabolism toward specific CYPs. To identify the isoform specificity, the query molecules need to fulfill at least one of the chemical features specified in a particular group as well as pharmacophoric features of corresponding group.

Group-1: (i) Presence of carboxylic acid (COOH) (ii) Presence of phenolic hydroxyl (iii) Presence of phenyl methylene with EWG (iv) Presence of sulfonamide (SO_2NH_2) (v) Presence of αβ-unsaturated carbonyl compound (vi) Presence of tetrazole and triazole functional units.

Group-2: Presence of (i) Aliphatic primary amines (ii) Aliphatic secondary amines (iii) Aliphatic tertiary amines

Group-3: (i) Presence of cyclohexyl, cyclopentyl, isopropyl, isobutyl structural unit (ii) Presence of aliphatic amines with halogens (iii) Absence of acidic structural environments specified in group-1 and basic structural environments depicted in group-2 (iv) Presence of three aromatic rings in the absence of group-1.

Decision tree development and validation to identify the isoform specificity

A databank was created to identify the isoform specificity of CYP substrates. It includes all the training sets and test sets of CYP substrates. Chemical feature based analysis of

substrates for identifying the specificity is mainly based on chemical functional groups. An attempt to identify the isoform specificity using chemical feature alone suffers from human judgmental errors. This is particularly true when there is more than one competing chemical feature in single substrates. The results of screening substrates using chemical features alone are given in Table 3. This approach results in many successful hits 27/30 for CYP2C9, 34/37 for CYP2D6 and 63/74 for CYP3A4. Though this is found to be satisfactory for the chosen cytochromes, many hits are obtained while cross screening, i.e., 17/37 for CYP2D6 vs. Group3.

The screening strategy based on pharmacophoric features alone can also be considered. The results from such a protocol are given in Table 4. The success rate in identifying substrate is relatively less for CYP2C9 (23/30), CYP2D6 (29/37) and more for CYP3A4 (71/74). However, specificity in selection is poor in this approach also, for example, CYP3A4 substrates against CYP2C9-hypo1 resulted in 45/74 successful hits. A combination of these two filters yielded better results. Table 5 shows the results of employing a decision tree (Fig. 3) which
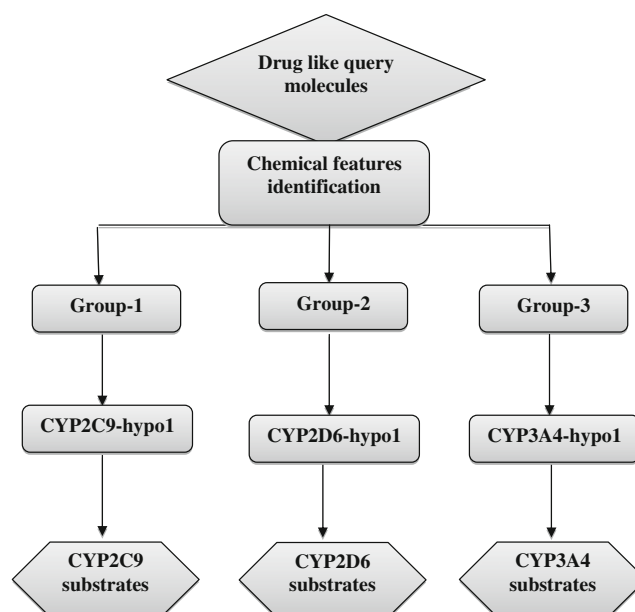
Fig. 3 Decision tree to predict the isoform specificity of CYP substrates

employs both (i) Chemical features and (ii) Pharmacophore hypotheses.

To validate the decision tree, a databank of compounds characterized by group-1 was subjected for the screening by CYP2C9-hypo1. Similarly all those compounds which were characterized by group-2 and group-3 respectively were subjected to screening by CYP2D6-hypo1 and CYP3A4-hypo1. CYP2C9-hypo1 predicted 23 (76.67%) of CYP2C9 compounds along with 3, 20 compounds respectively of CYP2D6 and CYP3A4 including their common substrates. CYP2D6-hypo1 predicted 29 (78.38%) of CYP2D6 compounds along with 2, 22 compounds respectively of CYP2C9 and CYP3A4 including their common substrates. CYP3A4-hypo1 predicted 61 (82%) of CYP3A4 compounds along with 7, 6 compounds respectively of CYP2C9 and CYP2D6 including their common substrates (Table 5). When the results were compared with experimental results the success rate of the protocol is higher in predicting the isoform specificity. Thus it can be concluded that the protocol suggested in the decision tree (Fig. 3) shall be useful in performing CYP isoform specificity analysis of any new drug toward its metabolism.

Conclusions

Three independent ligand based pharmacophore models were generated for CYP2C9, CYP2D6 and CYP3A4 using CATALYST software. Structurally diverse and highly active CYP substrates were employed in model building using common feature algorithm implemented in the HipHop module. The chosen models were validated using independent data sets of substrates. A protocol based on decision tree was developed to identify the isoform specificity associated with drug metabolism. This decision tree was also validated using a data bank of mixed CYP substrates. The decision tree predicted isoform specificity toward CYP2C9, CYP2D6 and CYP3A4 with higher accuracy. Validated protocol is useful in predicting the CYP isoform specificity in the metabolic profile of substrates metabolized by the cytochromes.

References

1. Guengerich FP, MacDonald JS (2007) Chem Res Toxicol 20:344–369
2. Ortiz de Montellano PR (2005) Cytochrome P450: structure, mechanism and biochemistry, 3rd edn. Kluwer, New York, pp 1–7
3. Rendic S, Carlo FJD (1997) Drug Metab Rev 29:413–580
4. Williams JA, Hyland R, Jones BC, Smith DA, Hurst S, Goosen TC, Peterkin V, Koup JR, Ball SE (2004) Drug Metab Dispos 32:1201–1208
5. Bazeley PS, Prithivi S, Struble CA, Povinelli RJ, Sem DS (2006) J Chem Inf Model 46:2698–2708
6. Guengerich FP (1999) Ann Rev Pharmacol Toxicol 39:1–17
7. Wester MR, Yano JK, Schoch GA, Yang C, Griffin KJ, Stout CD, Johnson EF (2004) J Biol Chem 279:35630–35637
8. Boelsterli UA (2003) Toxicol Appl Pharmacol 192:307–322
9. Margolis JM, O'Donnell JP, Mankowski DC, Ekins S, Obach RS (2000) Drug Metab Dispos 28:1187–1191
10. Rowland P, Blaney FE, Smyth MG, Jones JJ, Leydon VR, Oxbrow AK, Lewis CJ, Tennant MG, Modi S, Eggleston DS (2006) J Biol Chem 281:7614–7622
11. Terfloth L, Bienfait B, Gasteiger J (2007) J Chem Inf Model 47:1688–1701
12. Manga N, Duffy JC, Rowe PH, Cronin MT (2005) SAR QSAR Environ Res 16:43–61
13. Yap CW, Chen YZ (2005) J Chem Inf Model 45:982–992
14. Block JH, Henry DR (2008) J Comput Aided Mol Des 22:385–392
15. de Groot MJ, Alex AA, Jones BC (2002) J Med Chem 45:1983–1993
16. Jones BC, Hawksworth G, Horne VA, Newlands A, Morsman J, Tute MS, Smith DA (1996) Drug Metab Dispos 24:260–266
17. Mancy A, Broto P, Dijols S, Dansette PM, Mansuy D (1995) Biochemistry 34:10365–10375
18. Wolff T, Distlerath LM, Worthington MT, Groopman JD, Hammons GJ, Kadlubar FF, Prough RA, Martin MV, Guengerich FP (1985) Cancer Res 45:2116–2122
19. Meyer UA, Gut J, Kronbach T, Skoda C, Meier UT, Catin T, Dayer P (1986) Xenobiotica 16:449–464
20. Koymans L, Vermeulen NP, van Acker SA, Te Koppele JM, Heykants JJ, Lavrijsen K, Meuldermans W, Donné-Op KGM (1992) Chem Res Toxicol 5:211–219
21. Lewis DFV, Eddershaw PJ, Goldfarb PS, Tarbit MH (1996) Xenobiotica 26:1067–1086
22. Ekins S, Bravi G, Wikel JH, Wrighton SA (1999) J Pharmacol Exp Ther 291:424–433
23. Cruciani G, Carosati E, De Boeck B, Ethirajulu K, Mackie C, Howe T, Vianello R (2005) J Med Chem 48:6970–6979
24. Hennemann M, Friedl A, Lobell M, Keldenich J, Hillisch A, Clark T, Göller A (2009) Chem Med Chem 4:657–669
25. Boyer S, Hasselgren Arnby C, Carlsson L, Smith J, Stein V, Glen RC (2007) J Chem Inf Model 47:583–590
26. SYBYL version 6.9 (2007) Tripos Inc, 1699 South Hanley Road, St Louis, MO 63144, USA
27. CATALYST version 4.10 (2005) Accelrys Inc, San Diego, CA. http://www.accelrys.com
28. Adane L, Patel DS, Bharatam PV (2009) Chem Biol Drug Des 75:115–126
29. Sundriyal S, Sharma RK, Jain R, Bharatam PV (2008) J Mol Model 14:265–278
30. Williams PA, Cosme J, Ward A, Angove HC, Vinkovic DM, Jhoti H (2003) Nature 424:464–468
31. Ridderström M, Masimirembwa C, Trump-Kallmeyer S, Ahlefelt M, Otter C, Andersson TB (2000) Biochem Biophys Res Commun 270:983–987
32. Dickmann LJ, Locuson CW, Jones JP, Rettie AE (2004) Mol Pharmacol 65:842–850
33. Hanna IH, Krauser JA, Cai H, Kim MS, Guengerich FP (2001) J Biol Chem 276:39553–39561
34. Kirton SB, Kemp CA, Tomkinson NP, St Gallay S, Sutcliffe MJ (2002) Proteins 49:216–231
35. Paine MJI, McLaughlin LA, Flanagan JU, Kemp CA, Sutcliffe MJ, Roberts GCK, Wolf CR (2003) J Biol Chem 278:4021–4027

36. Fishelovitch D, Shaik S, Wolfson HJ, Nussinov R (2009) J Phys Chem B 113:13018–13025
37. Tang C, Shou M, Rodrigues AD (2000) Drug Metab Dispos 28:567–572
38. Iida I, Miyata A, Arai M, Hirota M, Akimoto M, Higuchi S, Kobayashi K, Chiba K (2004) Drug Metab Dispos 32:7–9
39. Tang C, Shou M, Mei Q, Rushmore TH, Rodrigues AD (2000) J Pharmacol Exp Ther 293:453–459
40. Bajpai M, Roskos LK, Shen DD, Levy RH (1996) Drug Metab Dispos 24:1401–1403
41. Ufer M, Kammerer B, Kahlich R, Kirchheiner J, Yasar U, Brockmöller J, Rane A (2004) Xenobiotica 34:847–859
42. Vickers AEM, Sinclair JR, Zollinger M, Heitz F, Glänzel U, Johanson L, Fischer V (1999) Drug Metab Dispos 27:1029–1038
43. Katoh M, Nakajima M, Shimada N, Yamazaki H, Yokoi T (2000) Eur J Clin Pharmacol 55:843–852
44. Venhorst J, Onderwater RCA, Meerman JHN, Commandeur JNM, Vermeulen NPE (2000) Drug Metab Dispos 28:1524–1532
45. Olesen OV, Linnet K (2000) Pharmacology 59:298–309
46. Matsumoto S, Yamazoe Y (2001) Br J Clin Pharmacol 51:133–142
47. Otton SV, Ball SE, Cheung SW, Inaba T, Rudolph RL, Sellers EM (1996) Br J Clin Pharmacol 41:149–156
48. Kim KA, Chung J, Jung DH, Park JY (2004) Eur J Clin Pharmacol 60:575–581
49. Olesen OV, Linnet K (1997) Pharmacology 55:235–243
50. Klees TM, Sheffels P, Dale O, Kharasch ED (2005) Drug Metab Dispos 33:303–311
51. Jacobsen W, Kuhn B, Soldner A, Kirchner G, Sewing KF, Kollman PA, Benet LZ, Christians U (2000) Drug Metab Dispos 28:1369–1378
52. Rakhit A, Pantze MP, Fettner S, Jones HM, Charoin JE, Riek M, Lum BL, Hamilton M (2008) Eur J Clin Pharmacol 64:31–41
53. Huang W, Lin YS, McConn DJ, Calamia JC, Totah RA, Isoherranen N, Glodowski M, Thummel KE (2004) Drug Metab Dispos 32:1434–1445
54. Jang GR, Wrighton SA, Benet LZ (1996) Biochem Pharmacol 52:753–761
55. Emoto C, Murase S, Sawada Y, Jones BC, Iwasaki K (2003) Drug Metab Pharmacokinet 18:287–295
56. Cameron MD, Wright J, Black CB, Ye N (2007) Drug Metab Dispos 35:1894–1902

ORIGINAL PAPER

Signal mass and Ca^{2+} kinetics in local calcium events: a modeling study

Irina Baran · Constanta Ganea · Raluca Ungureanu · Ioana Teodora Tofolean

Abstract We use a detailed modeling formalism based on numerical simulations of local calcium release events where the blurring of the image, the presence of diffusional barriers provided by large organelles situated close to the release site, as well as the variable position of the scan line with respect to the release site are taken into consideration. We have investigated the effect of the fluorescence noise fluctuations on the accuracy in computing the signal mass from linescan recordings and obtained a quantitative description of both the signal mass and the local increase in the free Ca^{2+} level as a function of the release current, the release duration and the orientation of the scan line, for three different levels of noise magnitudes. The model could provide a very good fit to a wide set of available experimental data regarding the signal mass of puffs visualized by fluorescence microscopy in the *Xenopus* oocyte loaded with 40 μM Oregon Green-1 in the absence of the calcium chelator EGTA. Numerical simulations also predict the amplitude and the kinetics of calcium signals evolving in the absence of the indicator, and indicate that sub-maximal activation of IP$_3$ receptors could produce in average levels of about 2 μM and 0.4 μM free Ca^{2+} close to a release site located in the animal or in the vegetal hemisphere, respectively, whereas the maximal levels reached in more rare events could be 11 μM and 4 μM, respectively.

Keywords Anisotropic diffusion · Calcium release · Linescan image · Numerical simulations · Signal mass

I. Baran (✉) · C. Ganea · R. Ungureanu · I. T. Tofolean
Department of Biophysics,
"Carol Davila" University of Medicine and Pharmacy,
8 Eroii Sanitari,
050474 Bucharest, Romania
e-mail: baran@ifin.nipne.ro

Introduction

Ca^{2+}, a ubiquitous second messenger involved in cellular signaling in cells ranging from bacteria to specialized neurons, is known to play a central role in the control of numerous physiological processes, including cell proliferation, gene transcription, secretion, contraction, fertilization or apoptotic/necrotic cell death [1–3]. A variety of stimuli can induce release of Ca^{2+} ions from the endoplasmic reticulum (ER) by activating Ca^{2+} channels which are organized in discrete clusters on the ER membrane [4–6]. An active release unit, comprised of one or more IP$_3$ receptors (IP$_3$Rs) which function as Ca^{2+} channels, can create a localized Ca^{2+} gradient in the surrounding space (a Ca^{2+} microdomain), which subsequently dissipates into the cytosol via Ca^{2+} diffusion and binding to fixed or mobile endogenous buffers. Space, time and intensity are three essential parameters that collectively shape local and global calcium signals, the building blocks of the calcium signaling tool kit that operates inside cells to generate transient Ca^{2+} signals, oscillations or waves. Ca^{2+} dynamics modeling has evolved from simple, low-dimensional models to highly complex approaches which generally consider the multi-facets processes of intracellular Ca^{2+} regulation, including the activity of the calcium release channel and its modulation by various agonists or antagonists, binding of Ca^{2+} ion to cytosolic mobile or fixed buffers, diffusion, Ca^{2+}/K$^+$ exchange inside the lumen of ER, or Ca^{2+} extrusion from the cytosol (reviewed in [7–9]). The problem may be further complicated by stochastic and noise fluctuations effects [10, 11] which may be coupled to relevant concentration gradients in the non-homogeneous cellular domain and thus can give rise to spontaneous Ca^{2+} oscillations [10–12]. The functional role of Ca^{2+} microdomains as potential sources for the generation of global Ca^{2+} signals is mediated by diffusive coupling

between Ca^{2+} release sites and by the specific feature of the Ca^{2+} release channels to activate upon binding of Ca^{2+} ion to the receptor molecules [8]. Since Ca^{2+} microdomains are stochastic in nature and they are not strongly coupled, the cellular ensemble of microdomains appears to behave stochastically rather than deterministically [10, 11]. This has significant physiological implications because such behavior can produce oscillatory dynamics for parameters outside the range exhibited by a purely deterministic system [10, 12]. Importantly, Perc et al. [13] have clearly established for the first time the stochastic nature of intracellular calcium oscillations directly from experimental data, and their analysis suggests that the stochastic nature of signals at cellular level becomes increasingly deterministic at the level of the organ.

A particular issue of Ca^{2+} dynamics regards the release kinetics at a local site and the corresponding dynamics of the associated microdomain. Extensive experimental [4–6, 14–21] and modeling [22–29] efforts in this area have collectively provided valuable data and information. However, essential parameters remain currently unknown, such as the number of releasing channels, release duration, the width of the channel cluster, and the size of the Ca^{2+} release flux. Puffs (large-amplitude signals) and blips (small-amplitude signals) are non-stereotypic local calcium events that arise in non-excitable cells, presumably through stochastic activation of one or more channels in the release unit [5, 6]. The time- and space-scales of such local Ca^{2+} signals which can develop over distances of several micrometers can be observed by fluorescence microscopy with the use of calcium-sensitive fluorophores. A particular method involves the construction of two-dimensional linescan images from dynamic recordings of the fluorescence emission of the dye along the direction of a laser beam which is spanning the cell. The data contained by such linescan images can provide relevant information on the dynamics of Ca^{2+} bound to the indicator. Thus, with the use of a minimal number of a priori assumptions, the release flux and duration can be estimated indirectly for every individual event [5, 6, 21, 22, 28, 29]. However, the in vivo release duration, the amount of liberated calcium or the number of conducting channels in a cluster during release remain currently uncertain mainly because our knowledge regarding the intracellular binding kinetics and availability of different endogenous or exogenous calcium buffers is rather limited. Recent modeling studies of calcium puffs in *Xenopus* oocytes have suggested an average release rate of 1.1 pA in the absence of the calcium chelator EGTA [28], and 0.34 pA [29] or 10 pA [25] in the presence of 300 μM EGTA, with estimated single channel currents on the order of 0.1 pA [28, 29] or 0.4 pA [25]. Strong experimental evidence has been provided for an IP_3R current ~0.05 pA in intact mammalian cells [16] or 0.12 pA in the outer nuclear envelope of DT40 cells [17], and comparable unitary currents of ~0.1–0.2 pA can be

extrapolated from data on the similar ryanodine receptor [8]. Moreover, a recent study provided from nuclear patch-clamp experiments the first measurement of the unitary current (~0.15 pA) through an open IP_3R channel in its native environment under physiological conditions [30]. In addition, the predictions for the average amount of liberated Ca^{2+} in vivo also present large variations, ranging from ~10 fC [29] to 130 fC [28] or 190 fC [25]. So, it is clear that variable experimental conditions related to the quantity and type of the calcium indicator used, or to the addition of a Ca^{2+} chelator, corroborated by differing binding scenarios inside the cytosol and the lumen of the endoplasmic reticulum, can lead to consistent differences between the modeling outputs [25, 27–29]. Moreover, these inherent differences can be further amplified when the effects of optical blurring, diffusional anisotropy and deviation of the sampling scan line from the release source are incorporated into the analysis [16, 23, 25–28].

In a previous paper [28] we analyzed a series of available experimental data of puff widths and fluorescence amplitudes obtained with the calcium indicator Oregon Green-1 in *Xenopus* oocytes in the absence of EGTA [5, 14, 15] and developed a model-based method for estimating the distributions of release currents and amounts of mobilized Ca^{2+}, as well as the distributions of the release site diameters in the two hemispheres of the oocyte. While the results therein provided a good representation of the data of puff fluorescence and spatial size, the experimental distribution reported for the signal mass [5] could not be replicated by our model. The concept of signal mass was originally introduced as a measure of the amount of calcium which is liberated during elementary release events [5]. According to its general definition, as will be described below in Methods, the signal mass quantifies the relative increase in the total amount of Ca^{2+} bound to the fluorescent dye. When applied to linescan experiments, the signal mass is usually calculated on the basis of two assumptions: that the calcium source is a point embedded in an isotropic environment, and that the fluorescence intensity corresponds to the local free calcium concentration at the confocal spot [5]. It has been generally assumed that the rate of rise of the signal mass is directly proportional to the underlying Ca^{2+} current and that its peak is proportional to the total Ca^{2+} flux [5, 21, 22]. However, calcium flux simulations indicate that significant deviations from this behavior may arise, mainly because Ca^{2+} is not at equilibrium with the indicator [22, 26]. The signal mass approach is of large interest because of its utility in providing an estimate for the Ca^{2+} release flux in both linescan recordings and wide-field microscopy. In principle, by differentiating the signal mass, the channel kinetics could be derived with sufficient accuracy to resolve channel openings and closings [22]. However, this procedure can be

affected by signal noise fluctuations which are inherent to the fluorescence recordings [22].

In this study we show that the release parameters predicted previously for the *Xenopus* oocyte (i.e., release duration, Ca^{2+} flux and size of IP_3R clusters) [28] are compatible with the experimental distribution of the signal mass [5] if the effect of noise fluctuations is incorporated in the derivation of signal mass. In this way we could obtain a very good representation of a large pool of signal mass data collected from 655 events observed in the *Xenopus* oocyte [5]. Consequently, this represents the first model, to the best of our knowledge, which is able to reproduce a wide set of data regarding the fluorescence amplitude, spatial size and signal mass of puffs visualized in the *Xenopus* oocyte with 40 μM Oregon Green-1 in the absence of EGTA, by taking into consideration the blurring of the image, the presence of diffusional barriers provided by large organelles situated close to the release site, the effect of noise, as well as the variable position of the scan line with respect to the release site. More importantly, this study provides for the first time a detailed quantitative analysis on how the signal mass depends on the total amount of mobilized Ca^{2+}, on the calcium flux and on the orientation of the scan line relative to the release site, and also indicates how much the signal mass is affected by different levels of noise. The current numerical simulations also predict the kinetics of the actual increase in the cytosolic free Ca^{2+} in a restricted domain around the release source both in the presence and in the absence of the fluorescent indicator, and indicate a consistent difference between the animal and the vegetal hemisphere of the oocyte.

The model

To simulate local Ca^{2+} release events we have considered one, two or three clusters of IP_3Rs placed on a tubule of the endoplasmic reticulum with radius $r_T = 200$ nm and calculated the time and space variation in the various species concentration following the synchronous Ca^{2+} release by 25–50 channels in each cluster. IP_3Rs in a cluster are arranged as a regular lattice on the surface of the ER tubule cylinder and the associated release fluxes are considered uniform through the ER membrane patch of the respective voxel [26–28].

The spatio-temporal evolution of each mobile specie concentration (Y_m) is calculated according to the reaction-diffusion equation:

$$\frac{\partial Y_m(r, \varphi, z, t)}{\partial t} = D_m \nabla^2 Y_m(r, \varphi, z, t) + F_{Y_m}(r, \varphi, z, t), \quad (1)$$

where D_m is the diffusion coefficient of the mobile specie, and F_{Y_m} is the corresponding buffer exchange term. The

mobile species are: Ca^{2+} and K^+ inside the ER, and Ca^{2+}, the endogenous mobile buffer and the exogenous calcium mobile buffer Oregon Green-1 in the cytosol. The standard buffer parameter values [26, 28] used in the current simulations are shown in Table 1. The cytosolic Ca^{2+} diffusion coefficient is set to 200 μm² s⁻¹. For each buffer, either mobile or immobile, the buffer exchange term can be written as:

$$F_X(r, \varphi, z, t) = -k_X^{on} \times [Ca^{2+}]_{cyt}(r, \varphi, z, t) \times [X] \quad (2)$$

$$(r, \varphi, z, t) + k_X^{off} \times [XCa](r, \varphi, z, t),$$

where X represents the buffer, k^{on} and k^{off} are on- and off-binding rate constants, $k_X^{off} = k_X^{on} K_{d,X}$, $K_{d,X}$ dissociation constant of buffer X, $[Ca^{2+}]_{cyt}$ cytosolic concentration of free Ca^{2+}, $[X]$ concentration of X and $[XCa]$ concentration of Ca^{2+}-bound X. Then the cytosolic buffering term for Ca^{2+} becomes:

$$F_{Ca,cyt}(r, \varphi, z, t) = \sum_X F_X(r, \varphi, z, t). \quad (3)$$

The evolution of the fixed buffer (B_{fix}) within the cytosol is determined from

$$\frac{d[B_{fix}](r, \varphi, z, t)}{dt} = F_{B_{fix}}(r, \varphi, z, t) \quad (4)$$

where $F_{B_{fix}}$ is defined by Eq. 2 with $X \equiv B_{fix}$. The conservation rule: $[X](r, \varphi, z, t) + [XCa](r, \varphi, z, t) = X_T$, where X_T is the total concentration of buffer X, is applied at every time step for each buffer.

Table 1 Standard buffer parameters used in numerical simulations

Buffer	Parameter	Value
Fixed endogenous buffer		
	Total concentration	300 μM
	Dissociation constant	10 μM
	On-rate constant	100 μM⁻¹s⁻¹
Mobile endogenous buffer		
	Total concentration	50 μM
	Dissociation constant	10 μM
	On-rate constant	100 μM⁻¹s⁻¹
	Diffusion coefficient	15 μm² s⁻¹
Calcium dye (Oregon Green-1)		
	Total concentration	40 μM
	Diffusion coefficient	20 μm² s⁻¹
	Dissociation constant	0.17 μM
	On-rate constant	75 μM⁻¹s⁻¹

It is assumed that Ca^{2+} and K^+ compete in binding to a common domain M of the glycoprotein matrix inside the ER, according to the Ca^{2+}/K^+ exchange model [12]:

$$MK \xrightleftharpoons[k_{MK}^{off}]{+K^+} M \xrightleftharpoons[k_{MCa}^{off}]{+Ca^{2+}} MCa$$
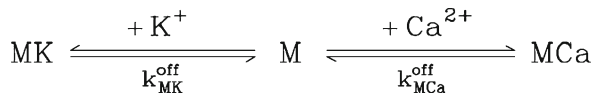
where ion binding is of first-order for calcium and h_K-order for potassium. Then we have:

$$\frac{d[MCa]}{dt} = k_{MCa}^{on} \times [M] \times [Ca^{2+}]_{lum} - k_{MCa}^{off} \times [MCa] \quad (5)$$

$$\frac{d[MK]}{dt} = k_{MK}^{on} \times [M] \times \left([K^+]_{lum}\right)^{h_K} - k_{MK}^{off} \times [MK] \quad (6)$$

$$\frac{\partial [Ca^{2+}]_{lum}}{\partial t} = D_{Ca}^{ER} \nabla^2 [Ca^{2+}]_{lum} - \frac{d[MCa]}{dt} \quad (7)$$

$$\frac{\partial [K^+]_{lum}}{\partial t} = D_K^{ER} \nabla^2 [K^+]_{lum} - h_K \times \frac{d[MK]}{dt} \quad (8)$$

$$[M] + [MCa] + [MK] = M_{max} \quad (9)$$

where $D_{Ca}^{ER} = 20\ \mu m^2\ s^{-1}$ and $D_K^{ER} = 50\ \mu m^2\ s^{-1}$ are the diffusion coefficients of Ca^{2+} and K^+ inside the ER, and M_{max} is the total concentration of Ca^{2+}/K^+ binding M-domains in the lumen. The Ca^{2+}/K^+ exchange is assumed to be fast: off-binding rate constants are $10\ s^{-1}$ for both Ca^{2+} and K^+. Other intralumenal parameters are as given previously [12].

Single channel ionic currents are considered of the form:

$$I(\varphi, z, t) = \min[I_N(\varphi, z, t), I_{sat}(\varphi, z, t)] \quad (10)$$

where

$$I_N(\varphi, z, t) = \frac{gRT}{nF} \times \ln\left[\frac{Y(r_T - \Delta r/2, \varphi, z, t)}{Y(r_T + \Delta r/2, \varphi, z, t)}\right] \quad (11)$$

is the local current driven by the Nernst potential, and

$$I_{sat}(\varphi, z, t) = 2\pi n F D_{Ca}^{ER} r_{cp} Y(r_T - \Delta r/2, \varphi, z, t) \times 10^{-3} \quad (12)$$

is the diffusion-limited saturating channel current [31]. Here the symbols are: $g=113$ pS, single channel conductance [4]; $n=2$, ion valence; $R=8.314$ J K^{-1} mol^{-1}, gas constant; $T=298$ K, temperature; $F=96500$ C mol^{-1}, Faraday constant; Y molar concentration of Ca^{2+} in the voxel facing the luminal/cytosolic face of the channel; $r_{cp}=12$ nm, the capture radius of the pore.

The reaction-diffusion equations are solved with an explicit finite difference formula in cylindrical coordinates with spatial steps $\Delta z = \Delta r = 100$ nm, and an angular step $\Delta \varphi = 9°$. The time step, ≤ 1.5 μs, is adjusted at each integration step to ensure the stability of the solution. The center of the tubule segment is considered the origin of the coordinate system, O. In all simulations we have assumed geometrical symmetry against the plane $z=0$. The ER tubule is oriented along the Oz axis. The simulated tubule segment is 9, 11 or 13 μm long if one, two or three clusters are considered. At the radial edge of the simulated cytosolic volume depth (5–7 μm from the tubule membrane) and at the z-edges fixed boundary conditions are applied according to the steady state values of the variables. At $r=0$ no-flux boundary conditions are applied. Neumann conditions at the ER/cytosol boundary are introduced according to the fluxes crossing each elementary ER membrane patch. We have also included uniformly distributed fluxes associated with the ER calcium pump [12]; however, we found that for the time scale of puffs the effect of the ER Ca^{2+}-ATPase can be neglected, consistent with other models [23, 24]. In addition, our model can also approximate the case in which the IP$_3$R cluster is located on a thin ER tubule (~50 nm diameter), in close proximity to a larger tubular structure.

The scan line is assumed to be parallel to the ER tubule at 350 nm away from the ER membrane; the central longitudinal row of channels in the cluster is viewed under a variable angle, α, against the axis Ox. The formation of the blurred optical image [23, 24] is converted to cylindrical coordinates. The point spread function of the confocal microscope is defined by a three-dimensional Gaussian function with lateral and axial confocal full-width at half-maximum $FWHM_{lat}$ and $FWHM_{ax}$, respectively. The fluorescence in a (r, φ, z) - voxel crossed by the scan line is calculated at every time step as the numerical approximation of the convolution integral:

$$F(r, \varphi, z, t) = \left(\sigma_{lat}^2 \sigma_{ax} \pi^{3/2}\right)^{-1} \iiint [DyeCa](r', \varphi', z', t) \quad (13)$$
$$\times \exp\left[-(r\cos\varphi - r'\cos\varphi')^2/\sigma_{lat}^2\right]$$
$$\times \exp\left[-(r\sin\varphi - r'\sin\varphi')^2/\sigma_{lat}^2\right]$$
$$\times \exp\left[-(z - z')^2/\sigma_{ax}^2\right] \times r'dr'd\varphi'dz'$$

where $\sigma = FWHM/[2(\ln 2)^{1/2}]$, $[DyeCa]$ is the concentration of Ca^{2+}-bound indicator, and $FWHM_{lat} = FWHM_{ax} = 400$ nm.

The scan line is divided in $2\Delta z = 200$ nm-length segments (pixels), to account for the experimental image resolution [5]. Fluorescence is spatially averaged over the couple of voxels overlapping a pixel, and the resulting

value defines the fluorescence intensity in the corresponding pixel at the respective time step. The center of the release event is estimated as the most likely symmetry center of the puff, and the fluorescence profile is averaged around that center over a distance of 600 nm along the scan line [5] to yield the fluorescence (F) signal associated to that event. F_0 denotes the resting fluorescence calculated in steady state, and $\Delta F = F - F_0$.

Since scan lines may be randomly oriented with respect to a certain cluster of IP$_3$Rs, different puff images are captured under various α angles which may vary between 0° and 180°. Then, the minimal distance from the voxel containing the nearby receptor to the pixel corresponding to the center of the puff falls between 250 and about 650 nm. A calcium event is considered to be generated by synchronous release through a certain number of clustered channels [25–29]. Some simulations were performed with two or three identical clusters of length L, separated by a side-to-side distance δ along the tubule, which release Ca^{2+} at the same time and the same rate [28]. Such cases are referred to as multiple (double or triple) events. The release duration (τ) is considered as an independent variable. The quantity of Ca^{2+} released by a cluster, Q, is estimated by numerical integration of $\int_0^\tau i\,dt$, where i is the total release current per cluster. In some sets of simulations we varied i, τ, and α between simulations. To vary the release current i, either the number of channels or the resting level of luminal Ca^{2+} was adjusted to obtain the required current amplitude [12]. The level of free Ca^{2+} in the lumen of the endoplasmic reticulum is generally considered to be between 100 and 700 μM [8, 30]. Accordingly, in our simulations the equilibrium level of luminal Ca^{2+} was considered in the range 150–750 μM, which produced single-channel Ca^{2+} currents within ~0.02–0.1 pA. Having in view that local luminal Ca^{2+} may be partially depleted during release, these figures are in very good agreement with recent measurements of the unitary current through an open IP$_3$R channel [30].

The fluorescence signal mass (σ_F) associated to each synthetic event is calculated by summing the fluorescence signals ($\Delta F/F_0$) along the scan line, multiplied by the volume of the spherical shell that includes the respective pixel. According to this algorithm, the observed center of the puff is considered the center of a sphere in a homogeneous medium, so that an isofluorescence spherical shell is associated at each moment to each pixel on the scan line [5]. We then calculate the numerical approximation of the integral $\sigma_F = \int (\Delta F/F_0)4\pi s^2 ds$, where s is the distance from the pixel to the center of the puff, measured along the scan line. σ_F is expressed signal-mass units (1 s.m.u. = $\Delta F/F_0 \times 10^{-15}$ l).

Consistent with the resolution of the imaging technique employed, we found that the image observed at 300–700 nm away from the release site is not affected by our discretization step. The large Ca^{2+} gradients that build up close to the channels are supplied by our numerical method in a smoothed form, as a spatial average over a distance of 100 nm. However, we found that this procedure does not affect the results obtained. We have checked this issue by using a different method with a variable radial step [32] which is increased progressively from 40 nm at the center of the tubule, to 300 nm at 7 μm away from the tubule, and obtained closely similar results.

To obtain the signal mass and free Ca^{2+} distributions we used the distributions of release durations, amounts of liberated Ca^{2+}, and release site diameters presented in Fig. 1. These distributions were shown [28] to reproduce accurately a series of experimental distributions of the fluorescence amplitudes and puff widths observed in *Xenopus* oocytes with Oregon Green [5, 14, 15] and are representative for a total number of 857 and 281 events recorded in the animal and in the vegetal hemisphere, respectively. Because different sets of data of fluorescence amplitudes and puff widths were employed in the analysis, which were collected from a variable number of events observed under similar conditions [5, 14, 15], the overall distribution of a given quantity displays a double-peak profile (Fig. 1), with a dominant maximum originating mainly from singular events and a smaller peak produced by the contribution of multiple release sites (i.e., two or three adjacent IP$_3$R clusters which are activated simultaneously and thus are observed as a single release site). Marked differences appear in the distributions of Ca^{2+} amounts in the vegetal hemisphere (Fig. 1b), where the small number of available experimental observations and the relatively large variations between individual sets of data produce a pronounced secondary peak. Shortly, for each distinct set of experimental data, a certain distribution of Qs released by individual sites was obtained [28]. Then the overall distributions were estimated by weighting each Q-distribution by its relative fraction of observations. The general form of the Gaussian distribution of Ca^{2+} amounts corresponding to a certain type of release sites was considered as follows:

$$\mathrm{d}n/\mathrm{d}Q \sim \exp\left[-\left(Q - \overline{Q}\right)^2/2/\sigma_Q{}^2\right]/\sigma_Q \qquad (14)$$

where n denotes generically the number of events. The standard values of the corresponding parameters $\overline{Q}$ and σ_Q are collected in Table 2 together with the relative frequency of different event classes. For each pair $(\overline{Q}, \sigma_Q)$ the Gaussian distribution was computed separately. All these distributions were then weighted by their relative frequency and summed up to obtain the curves presented in Fig. 1b.
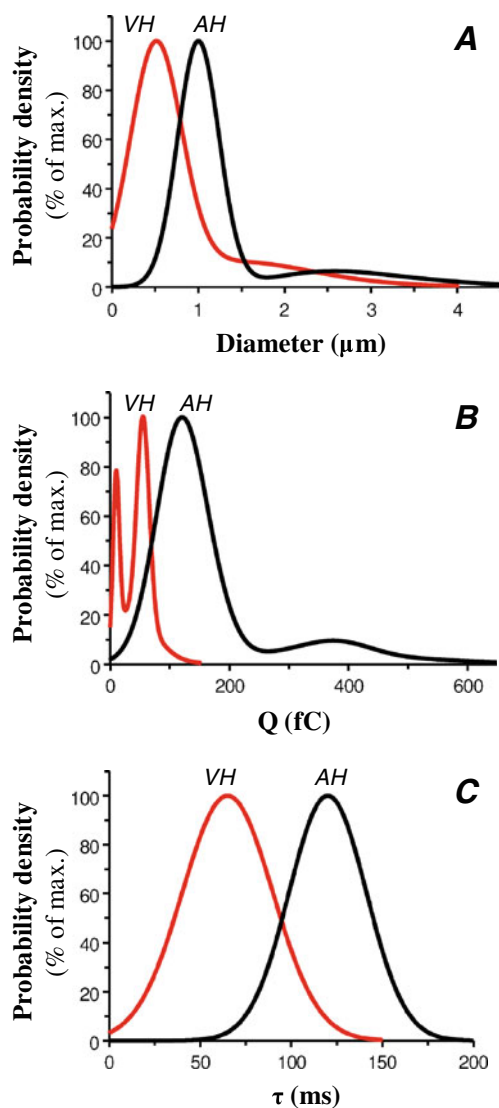
Fig. 1 Distributions of release site diameters (**a**), quantities of liberated Ca^{2+} (**b**) and release durations (**c**) assumed in the model for individual puffs evolving in the animal ("AH") or the vegetal ("VH") hemisphere of the oocyte. Each distribution is obtained as a linear combination of Gaussian functions, with parameters values provided in Table 2

The normal distribution of cluster diameters (L) was considered of the form:

$$dn/dL \sim \exp\left[-\left(L - \overline{L}\right)^2/2/\sigma_L^2\right]/\sigma_L \tag{15}$$

For simplicity, the inter-cluster distance (δ) in the animal/vegetal hemisphere was set to 0.39 μm and 0.4 μm, respectively [28]. The diameter of a release site was considered to be equal to L, $2L + \delta$ and $3L + \delta$ for singular, double and triple events, respectively. After setting the parameters values and relative frequencies (Table 3), the resulting distribution of release site diameters (Fig. 1a) was

calculated similarly to the Q-distribution. The release durations (Fig. 1c) were considered to follow a Gaussian distribution of the form:

$$dn/d\tau \sim \exp\left[-(\tau - \overline{\tau})^2/2/\sigma_\tau^2\right]/\sigma_\tau \tag{16}$$

with $\overline{\tau} = 120$ ms and $\sigma_\tau = 21.2$ ms in the animal hemisphere and $\overline{\tau} = 65$ ms and $\sigma_\tau = 24.8$ ms in the vegetal hemisphere.

To derive the signal mass distribution we divided sufficiently wide intervals for Q, τ, and α in discrete steps of 0.25 fC, 0.2 ms and 5°, respectively. For each of these discretized variables we constructed independent distributions for Q (as described above), τ (Eq. 16), and α (a uniform distribution, i.e., α may take any value from 0° to 180°, with equal numbers of observations among all the discrete intervals). Then, for each individual set (Q,τ,α) we determine $i = Q/\tau$ and calculate the signal mass according to the leading equation (described in Section 3.1). The running numerical code unfolds the three concatenated distributions and counts at each iteration the events described by a certain signal mass value found at increments of 0.1 s.m.u. Finally, we construct the resulting distribution by representing the number of counted events as a function of the signal mass.

Results

Signal mass is largely affected by noise

First we investigated the effect of fluorescence noise and space discretization on the accuracy in computing the signal mass from linescan records. We found in our numerical simulations that the estimation of the amount of released Ca^{2+}, especially for small-amplitude events, can be accompanied by large variations due to intrinsic limits and uncertainties related to: 1) averaging fluorescence signals over discrete voxels of relatively large dimensions, 2) evaluation of the puff center location and its positioning in the appropriate voxel, 3) inherent noise fluctuations of the fluorescence signals, associated with both light scattering by inhomogeneous cellular composites and optical blurring by the microscope, and 4) use of the third power of the estimated distance to the apparent center of the puff. Each of these factors may introduce by itself major uncertainties and therefore produce appreciable errors in calculation of the signal mass [5, 18, 21, 22, 29].

Sun et al. [5] have obtained, in the animal hemisphere of the *Xenopus* oocyte, a signal mass distribution which extends up to about 100 s.m.u. and exhibits a prominent exponential component with an average value of 19.4 s.m.u. We examined whether our model could explain the experimental distribution of the signal mass. First we have

Table 2 Standard parameters of the Gaussian distributions for released Ca^{2+} amounts in the two hemispheres of the oocyte (derived from [28])

	$\overline{Q}$ (fC) per event	σ_Q (fC) per event	Relative weight (% of total)
Animal hemisphere	120[a]	42.4[a]	79.51[a]
	190[a]	28.3[a]	2.28[a]
	240[b]	84.8[b]	2.52[b]
	360[c]	127.3[c]	8.40[c]
	380[b]	56.6[b]	6.08[b]
	570[c]	84.9[c]	1.22[c]
Vegetal hemisphere	10[a]	5.3[a]	19.64[a]
	20[b]	10.6[b]	4.91[b]
	30[a]	14.1[a]	4.78[a]
	55[a]	10.6[a]	50.18[a]
	60[b]	28.2[b]	17.07[b]
	90[c]	42.4[c]	3.41[c]

[a] single events (Ca^{2+} released by individual IP_3R clusters)
[b] double events (Ca^{2+} released by 2 coupled IP_3R clusters)
[c] triple events (Ca^{2+} released by 3 coupled IP_3R clusters)

obtained a quantitative description of the signal mass (denoted here as σ_0) as a function of i, τ and α. By analyzing a wide set of simulation data, we could construct the heuristic function:

$$\sigma_0(i, \tau, \alpha) = k_0 i \tau / (1 + i/i_0)/(1 + \tau/\tau_1)$$
$$\times \exp[-\alpha(1 + \tau_2/\tau)/\alpha_0], i > 0, \tau > 0, \quad (17)$$

which provided a good agreement with the numerical data (see below). In Eq. 17, k_0=0.69 s.m.u./fC, i_0=65 pA, τ_1= 490 ms, τ_2=15 ms, and α_0=615°. However, based on the description offered by Eq. 17, we could not produce a reasonable fit to the experimental data of Sun et al. [5] (discussed below). This discrepancy would suggest, as a possible explanation, that the experimental signal mass values may have been underestimated due to reasons presented above.

Consequently, we investigated how the signal mass distribution changes when the fluorescence signal cannot be entirely detected along the scan line. The relation in Eq. 17 has been obtained under the assumption that noise fluctuations do not distort the maximal value of the signal mass. However, if, for example, we calculate the signal mass by taking into account only those voxels along the scan line in which $\Delta F/F_0 \geq \Delta F_n/F_0$, we obtain a lower value

of the signal mass. We found that the experimental distribution of the signal mass can be reproduced if we consider $\Delta F_n/F_0$=0.2 (discussed below). Here we present some results obtained with two different levels of noise, $\Delta F_n/F_0$=0.1 and $\Delta F_n/F_0$=0.2, respectively, and we designate the corresponding values of the signal mass as σ_1 and σ_2, respectively. The example presented in Fig. 2, which is representative for a typical puff in the animal hemisphere of the oocyte (Q=132 fC, τ=120 ms), illustrates how large the difference between the three representations of the signal mass may be.

We were interested to see how the apparent signal mass depends on the quantity of released Ca^{2+} under a broad range of physiological release conditions and to obtain corresponding functions to describe this dependence (note that these functions are uniquely defined irrespective of the distributions of release quantities and durations). Thus, in Fig. 3 we collect the values of the signal mass obtained from simulations with release currents $i \leq 5$ pA and release durations $\tau \leq 300$ ms. For each simulation, which was ran with a predetermined set of parameters, namely τ, the number of channels, the cluster size (0.5 $\mu m \leq L \leq 1.5$ μm), and the resting luminal $[Ca^{2+}]$ (as described in Methods), the values of Q and i were computed. Then, for each individual simulation, different scanning angles ($\alpha \leq 180°$)

Table 3 Standard parameters of the Gaussian distributions for release site diameters in the two hemispheres of the oocyte (derived from [28])

	$\overline{L}$ (μm)	σ_L (μm)	Apparent diameter	Relative weight (% of total)
Animal hemisphere	1	0.24	L	22.00
	1	0.24	$2L + \delta$	18.00
	1	0.24	$3L + 2\delta$	60.00
Vegetal hemisphere	0.47	0.33	L	39.42
	0.47	0.33	$2L + \delta$	9.86
	0.55	0.26	L	9.59
	0.55	0.26	$2L + \delta$	34.26
	0.55	0.26	$3L + 2\delta$	6.84
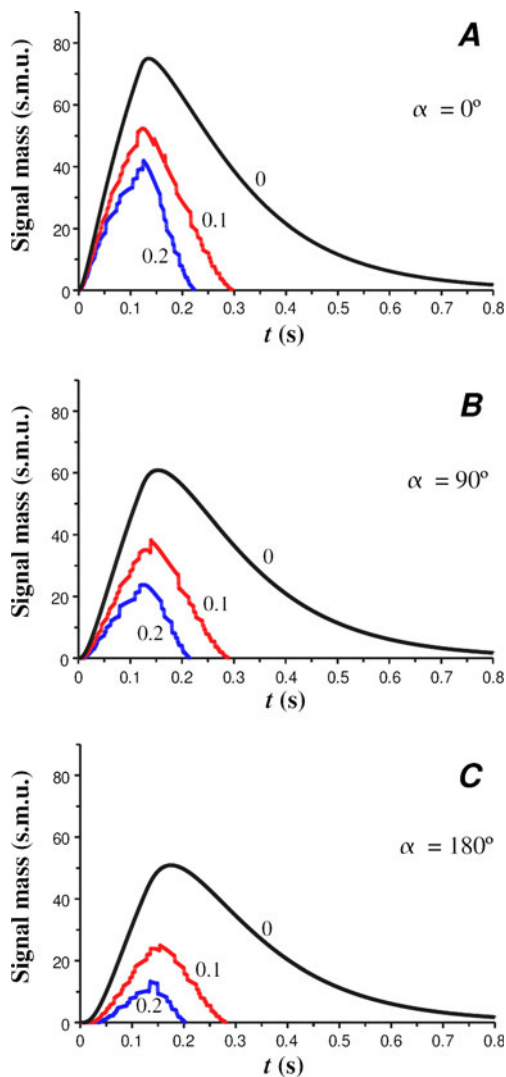
Fig. 2 Time course of the signal mass during a typical puff in the animal hemisphere (total amount of liberated Ca^{2+} 132 fC, release duration 120 ms), computed for fluorescence noise levels of 0, 0.1 or 0.2, as indicated. The orientation angle of the scan line is 0° (**a**), 90° (**b**) or 180° (**c**)

were considered for the computation of the signal mass variants, σ_0, σ_1 and σ_2. As expected, $\sigma_2 < \sigma_1 < \sigma_0$. The effect of the noise weakens at increasing Q, so that both σ_1 and σ_2 approach σ_0 when large quantities of Ca^{2+} are released. However, it appears that for typical Ca^{2+} release amounts and durations, i.e., $Q < 650$ fC and $\tau < 200$ ms (Fig. 1 here, [28]), the signal mass can exhibit tremendous variability depending on the quantity of liberated Ca^{2+}, release duration, orientation of the scan line and the magnitude of noise fluctuations. The error in signal mass estimation from noisy records appears to be exacerbated in slow release and distal sampling events, which correspond to the lower values plotted in graphs from Fig. 3b or c.

By analyzing the variation of σ_0, σ_1, σ_2 and the corresponding quantities $k_0 = \sigma_0/Q$, $k_1 = \sigma_1/Q$, $k_2 = \sigma_2/Q$ over

a large pool of simulation data, we obtained the following heuristic functions which could provide a good fit to all the simulation data:

$$\sigma_1(i, \tau, \alpha) = k_n i \tau / \left\{ 1 + \left[(i_1(\tau, \alpha)/i)^{h(\tau, \alpha)} \right] \right\} \tag{18}$$

$$/ \left\{ 1 + \left[(\tau_{n1}(i, \alpha)/\tau)^{h_1} \right] \right\}$$

$$/ \left\{ 1 + \left[(\tau/\tau_{n2}(\alpha))^{h_2} \right] \right\}, i > 0, \tau > 0,$$

where

$$i_1(\tau, \alpha) = i_{10}/(1 + \tau/\tau_i) \times \exp(\alpha/\alpha_i) \tag{19}$$

$$h(\tau, \alpha) = h_0/(1 + \tau/\tau_h) \times \exp(\alpha/\alpha_h) \tag{20}$$

$$\tau_{n1}(i, \alpha) = \tau_{n10} \times \exp(-i/i_{n1}) \times \exp(\alpha/\alpha_{n1}) \tag{21}$$

$$\tau_{n2}(\alpha) = \tau_{n20} \times \exp(\alpha/\alpha_{n2}) \tag{22}$$

with $k_n = 0.6$ s.m.u./fC, $i_{10} = 0.31$ pA, $\tau_i = 430$ ms, $\alpha_i = 103°$, $h_0 = 2$, $\tau_h = 400$ ms, $\alpha_h = 330°$, $\tau_{n10} = 4.4$ ms, $i_{n1} = 4.1$ pA, $\alpha_{n1} = 42°$, $h_1 = 0.71$, $\tau_{n20} = 440$ ms, $\alpha_{n2} = 360°$, $h_2 = 1.27$, and

$$\sigma_2(i, \tau, \alpha) = \sigma_1(i, \tau, \alpha) / \left\{ 1 + \left[(i_2(\tau, \alpha)/i)^{H(\alpha)} \right] \right\}, i > 0, \tau > 0 \tag{23}$$

where

$$H(\alpha) = H_0 [1 + (\alpha/\alpha_{H2})^r] \tag{24}$$

$$i_2(\tau, \alpha) = i_{20} \times \exp[\alpha/(\alpha_{i2} + a\tau)] \tag{25}$$

with $H_0 = 0.72$, $\alpha_{h2} = 165°$, $r = 6$, $i_{20} = 0.3$ pA, $\alpha_{i2} = 45°$, $a = 1°$ ms^{-1}.

In Figs. 4, 5 and 6 we illustrate some intermediary steps in obtaining these functions and the agreement with the simulation data. Here, the simulations were generally performed as described above. Hence, Fig. 4 presents an example of the signal mass dependence on the release current at fixed release duration (20 ms) for three different angles of the scan line, together with the corresponding fit of the simulation data to Eqs. 17–25, whereas in Fig. 5 we present the variation of k with the release duration for a fixed current of 2.4 pA. In Fig. 6 the simulation data are compared with the values calculated with Eqs. 17–25 for some representative scanning angles.

After obtaining the quantitative description of the signal mass presented above, we can now compute the signal mass distribution generated by a given set of distributions of release quantities Q and durations τ (as described in
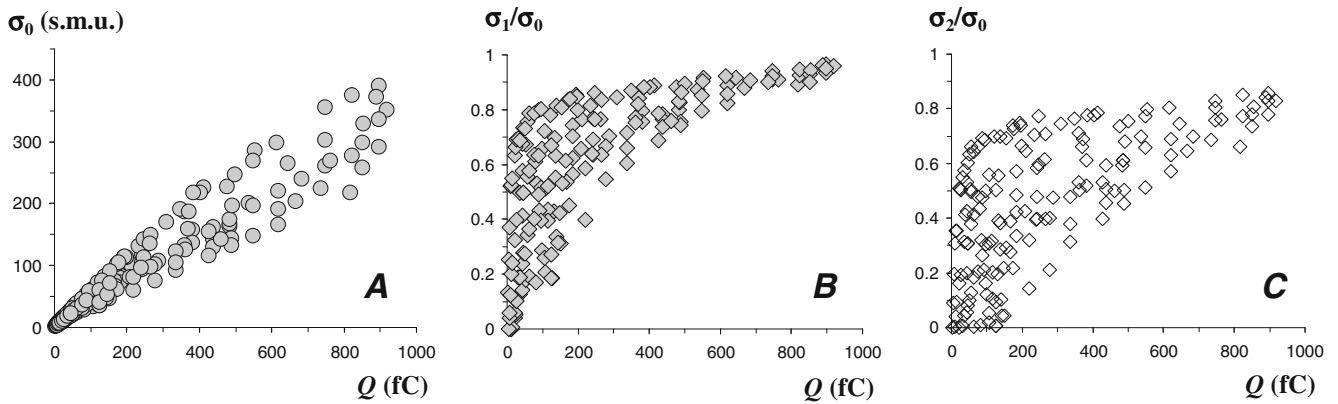
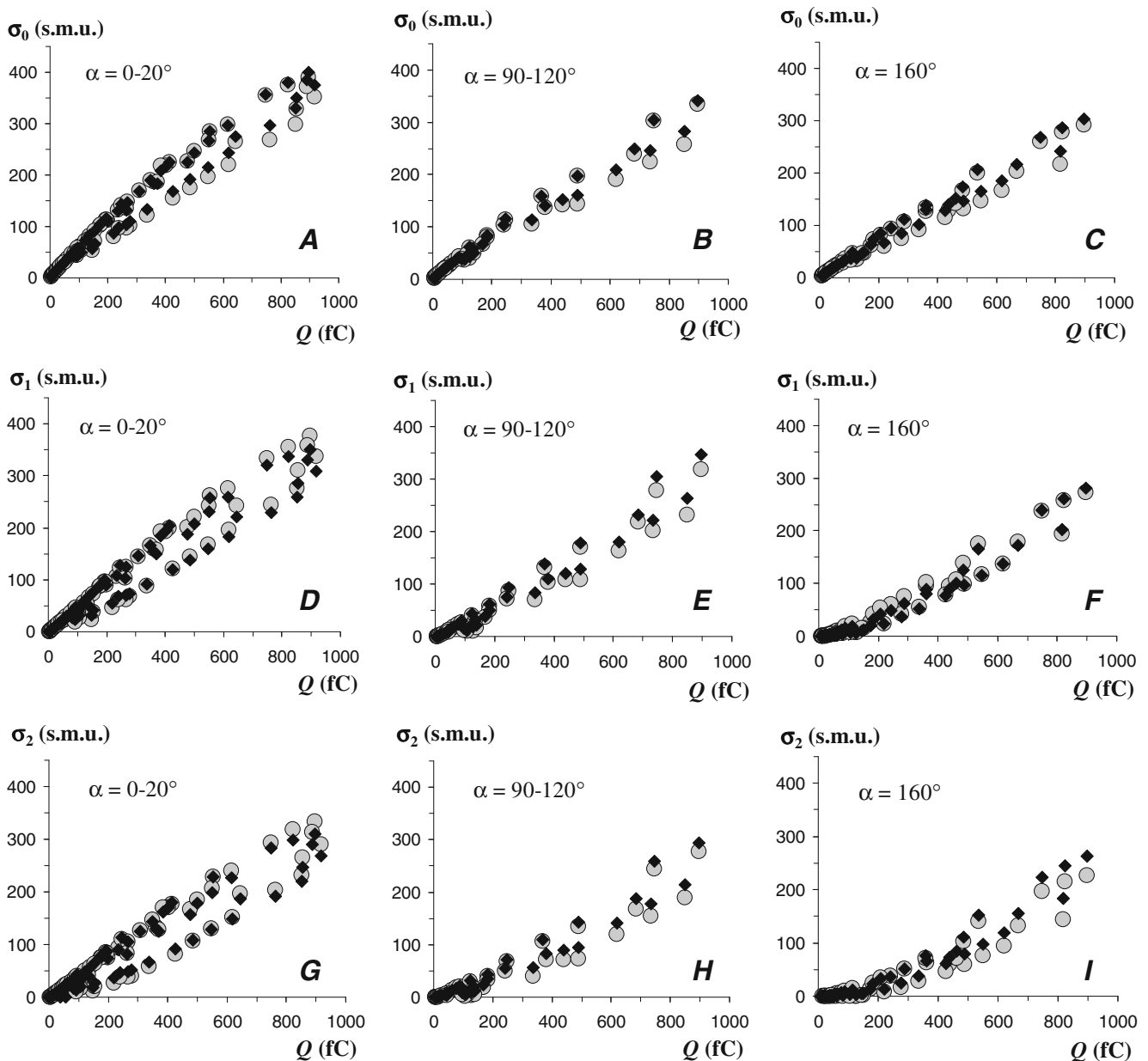**Fig. 3** Dependence of the signal mass on the quantity of liberated Ca^{2+} (Q), obtained from numerical simulations with release currents ≤5 pA, release durations ≤300 ms and scanning angles ≤180°. In (**a**) the signal mass is calculated without taking into consideration the effect of noise fluctuations (σ_0). In (**b**) and (**c**) the signal mass is computed by assuming a noise level of 0.1 (σ_1) and 0.2 (σ_2), respectively, and the ratio σ_1/σ_0 or σ_2/σ_0 is plotted against Q

Methods) and confront it to the experimental data (Fig. 7). Compared with other calcium indicators such as Fluo-4 dextran, Oregon Green provides a narrower range of fluorescence amplitudes and noisier signals, which make puffs harder to detect [5, 25, 29]. So, very low-amplitude signals emitted by Oregon Green can be missed in recordings and therefore can affect noticeably the signal mass value. Our investigations indicate that a noise level $\Delta F_n/F_0$ =0.2 in fluorescence recordings obtained with Oregon Green can provide an excellent agreement with the experimental distribution of the signal mass (Fig. 7). This figure ($\Delta F_n/F_0$=0.2) represents≈4% of the maximal fluorescence, $F_{max}/F_0≈5$, which is currently obtained with the Ca^{2+}-saturated form of this dye [5]. Therefore, our results appear to be realistic and, moreover, provide a consistent explanation for the experimental distributions of the fluorescence signals, puff widths and signal masses

detected in the animal hemisphere of the *Xenopus* oocyte. In Fig. 7 we also present the signal mass distributions computed for all the three levels of fluorescence noise in both hemispheres of the oocyte. For $\Delta F_n/F_0$=0.2, a similar shape of the signal mass distribution is obtained in the vegetal hemisphere; however, a 2–3 fold reduction in the signal mass values is observed.

Free Ca^{2+} levels during puffs

We were interested in the amplitude of local Ca^{2+} transients during puffs. To this end, we computed the maximal concentration of free Ca^{2+} reached in the front of the releasing IP_3R cluster ($\alpha=0°$), at 250–450 nm away from the center of the cluster, $[Ca^{2+}]_m$. We also determined the Ca^{2+} increase in the simulation domain surrounding the ER tubule on a space scale of 5–10 μm (volume V_s=783.74 fl),
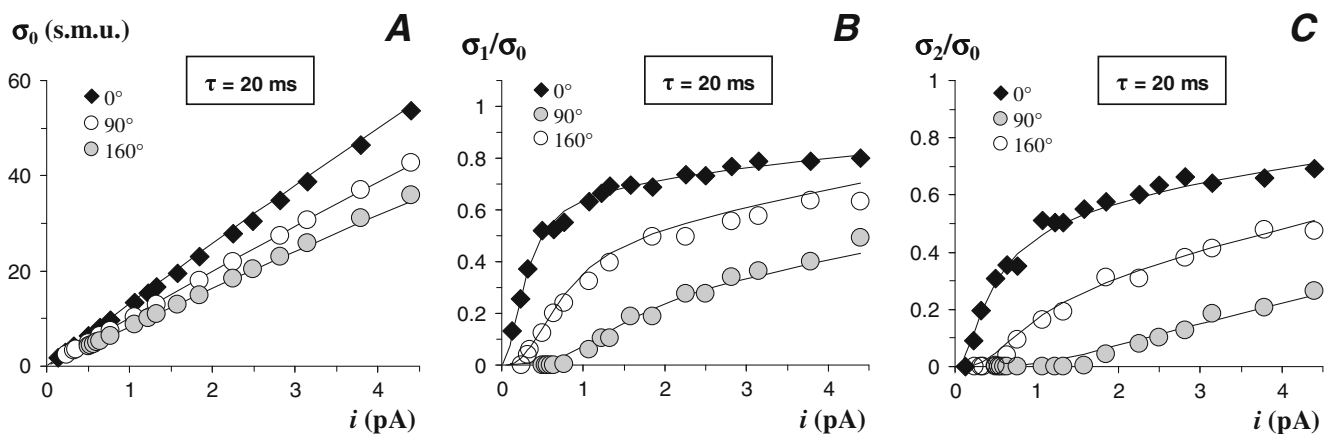
Fig. 4 Variation of the signal mass with the release current (i) for a fixed release duration (τ). Simulation data are obtained at three different angles of the scan line, as indicated in the legend. The best fit of Eqs. 17–25 to the numerical data is shown (curves). The noise level is 0 for σ_0, 0.1 for σ_1 and 0.2 for σ_2. σ_0 (**a**), σ_1/σ_0 (**b**) or σ_2/σ_0 (**c**) are plotted against i
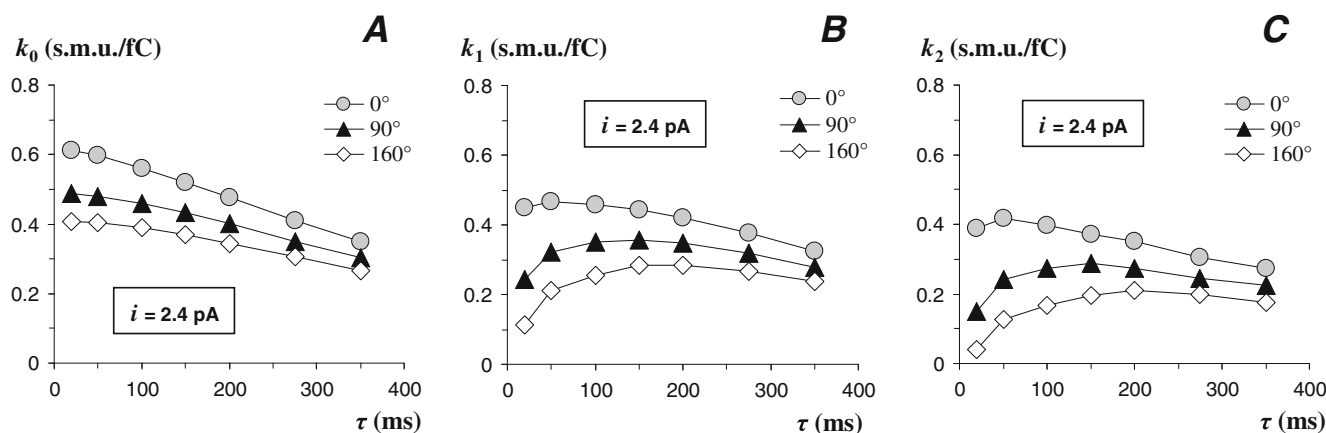
Fig. 5 Variation of the quantities $k_0 = \sigma_0/Q$ (**a**), $k_1 = \sigma_1/Q$ (**b**) and $k_2 = \sigma_2/Q$ (**c**), with the release duration τ, for a fixed release current i. The data are obtained from numerical simulations at three different angles of the scan line, indicated in the legend

which is denoted here as $[Ca^{2+}]_s$. By analyzing the simulation data we could select the following functions that fit well these data (Fig. 8):

$$\Delta[Ca^{2+}]_m(i,\tau) = [Ca^{2+}]_0 \times a_m/\{1 + [i_c/i/(1 + \tau/\tau_c)]^u\}, i > 0, \tau > 0,$$

(26)

with $a_m = 570$, $i_c = 8.0$ pA, $\tau_c = 405$ ms, $u = 1.78$, and

$$\Delta[Ca^{2+}]_s(i,\tau) = \Delta[Ca^{2+}]_m(i,\tau) \times a_s/[1 + \tau_s/\tau/(1 + i/i_s)], i > 0, \tau > 0,$$

(27)

with $a_s = 0.0026$, $\tau_s = 160$ ms, and $i_s = 0.85$ pA.

According to the results presented in Fig. 8a, $[Ca^{2+}]_m$ increases with the release current and can reach values ~250 higher than the resting level (i.e., ~8.5 μM) in the presence of 40 μM Oregon Green-1. The increase in free Ca^{2+} measured on a wider scale is generally damped by buffering and diffusion, so that $[Ca^{2+}]_s$ exhibits modest variations, up to about one half of the resting level. As expected, there is a direct correlation between $[Ca^{2+}]_m$ and $[Ca^{2+}]_s$ (Fig. 8b), which is also reflected by Eqs. 26–27.

The functions given in Eqs. 26–27 enable us to predict the distributions of these two quantitative measures of Ca^{2+} transients in both the hemispheres of the oocyte (Fig. 9), by employing a procedure similar to that used for the computation of the signal mass distributions, with the exception that the angular distribution was not taken into account. According to these calculations, the most frequent events in the animal hemisphere appear to produce a ≈ 25-fold increase in $[Ca^{2+}]_m$ above the resting level (i.e., 850 nM), whereas in the vegetal hemisphere events with very low amplitude (< 200 nM Ca^{2+} near the release site) appear to prevail. For a medium level of IP_3 stimulation as used in experiments [5], the free Ca^{2+} increase in the cytosolic area surrounding the cluster appears to be extremely reduced (Fig. 9b). However, it should be noted that these values

have been obtained under the condition that an exogenous buffer is present (more aspects will be presented in the Discussion). In the next section we investigate how a typical calcium puff would evolve in the absence of the calcium indicator.

Kinetics of an average puff in the presence/absence of the dye

In a previous paper [28] we estimated that an average puff evoked by medium IP_3-stimulation in the animal hemisphere of the oocyte is characterized by $Q = 132$ fC and $\tau = 120$ ms. Here we present the associated kinetic profiles of the fluorescence signal detected at different orientation angles of the scan line, together with the underlying time course of the maximal concentration of free calcium reached on the scan line, $[Ca^{2+}]_{max}$ (Fig. 10). The inset in Fig. 10 also shows the time evolution of the total release current that generates this average puff. One can notice a much larger variation in the effective level of free Ca^{2+} when the scan line is rotated around the ER tubule, as compared with the actual amplitude of the fluorescence signal emitted by the calcium-sensitive dye. As discussed previously [26, 27], there are two main reasons for this difference: 1) Ca^{2+} is not at equilibrium with the dye since binding is relatively slow, and 2) the recorded fluorescence data incorporate the blurring effect originating from more distant areas around the point where the signal is detected. Consequently, the derived relative increase is 38-fold for $[Ca^{2+}]_{max}$ but only 1.8-fold for F (both figures are relative to resting level) when the puff is observed in the front of the cluster. These values decrease 38 and 4 times, respectively, when the puff is detected at 180°. The predicted maximal concentration levels of free calcium along the scan line are 1.3 μM, 217 nM and 69 nM under scanning angles of 0°, 90° and 180°, respectively.

Fig. 6 Variation of the signal mass with the quantity of liberated Ca^{2+} (Q), obtained from numerical simulations with release currents≤5 pA, release durations≤300 ms and scanning angles α indicated in each panel. The noise level is 0 for σ_0 (**a-c**), 0.1 for σ_1 (**d-f**) and 0.2 for σ_2 (**g-i**). *Circles*, simulation data; *diamonds*, values calculated from Eqs. 17–25

We investigated the relative contribution of various buffers to the shape of the Ca^{2+} transient which develops in the cytosolic domain surrounding the release unit (Fig. 11a) for an average puff detected with 40 μM Oregon Green-1. We calculated the corresponding increase in the level of Ca^{2+} bound to the dye (OG) and to the endogenous mobile (MB) or fixed (FB) buffers, as well as the level of total free + bound Ca^{2+} (total). The total Ca^{2+} increases steadily during the release interval, up to a maximal value of 350.9 nM, after which decays to the resting level with a half-time of 170 ms.

A large part of the released Ca^{2+} spreads further away to distal sites through clearance processes inside the cytosol. Thus, we calculated that a maximal concentration of 875.6 nM should be reached if all the liberated calcium remained confined into the computation domain (volume V_s= 783.74 fl). A large part (73%) of the total Ca^{2+} appears to be bound to the indicator, whereas the remaining calcium in bound form is distributed between FB (23%) and MB (4%), respectively. Finally, there is a negligible increase in the level of free Ca^{2+} (2.9 nM at its peak).
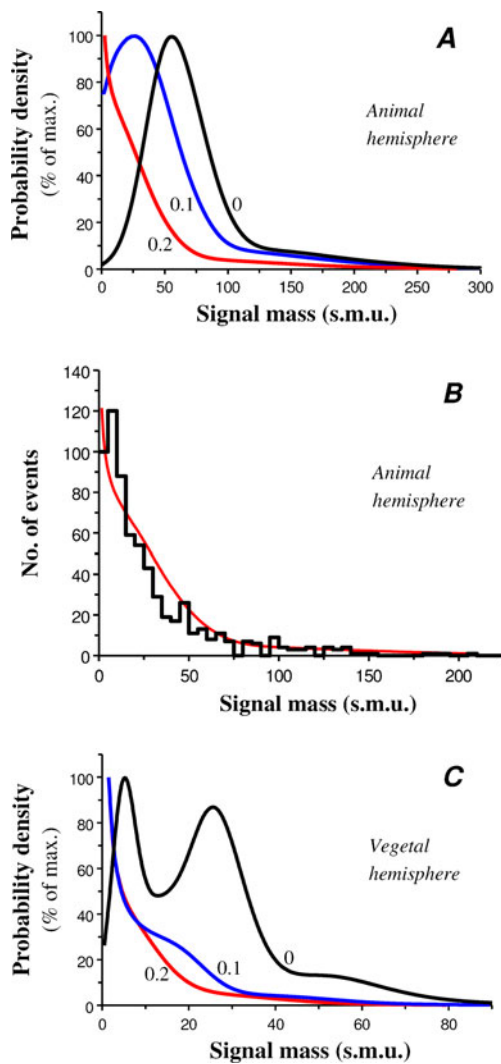
Fig. 7 Signal mass distribution. (a) The distribution of the signal mass in the animal hemisphere of the oocyte was computed for three noise levels indicated on each curve. Event frequency is normalized to its maximal value. (b) The experimental histogram obtained by Sun et al. (1998) in the animal hemisphere is compared with the theoretical distribution calculated for a noise level of 0.2 (continuous line). (c) Similar to (a), but for the vegetal hemisphere

In the absence of the indicator, simulations with the same release parameters as before (Q=132 fC and τ= 120 ms) predict that the rise phase of the total Ca^{2+} is virtually identical to that found in the previous situation, indicating a peak of 364.7 nM reached at the end of the release period (Fig. 11b). However, the decay phase appears to be slower, with a half-decay time of 432 ms, which is 2.5-fold lower than obtained in the presence of the dye. Our data suggest that the major part of Ca^{2+} binds to the endogenous fixed buffer (83% of total) while a smaller part is bound to the mobile buffer (14% of total). Again, one can notice that a major effect of Oregon Green-1 is to accelerate the relaxation phase following the release period, which can be explained by the high affinity of this dye. Another

important effect of Oregon Green-1 is the reduction in the level of free Ca^{2+}. Hence, the maximal increase in the concentration of free calcium evaluated in the entire computation domain decreases 3.6 times in the presence of the dye (Fig. 11c). Similarly, the maximal increase in the concentration of free Ca^{2+} close to the release site is reduced 1.66 times, from 2.16 μM in the absence of Oregon Green to 1.3 μM in the presence of 40 μM OG (Fig. 11d).

Discussion

Our numerical simulations indicate that the calcium-sensitive dye Oregon Green can affect considerably the kinetics and spatial distribution of Ca^{2+} released from the endoplasmic reticulum into the cytosol, which also implies that the release activity of the IP_3 receptors may differ in the presence of the indicator. Consequently, the very distributions of the amounts of released calcium and release currents may change when they are probed with calcium-sensitive indicators.

An interesting result in this study is that during a typical release event (120 ms release duration) only 40% of the released Ca^{2+} remains inside a cytosolic domain of≈800 fl around the release site, irrespective of the presence or the absence of Oregon Green, indicating that within 120 ms more than 50% of the liberated calcium is passed on by buffers or spreads freely by diffusion to distances > 5 μm from the release site. This could have implications in the generation of global signals when higher levels of IP_3 develop inside the cell. Recent measurements performed with the use of a multi-focal fluorescence microscopy system of high resolution indicated that elemental release sites in the *Xenopus* oocyte are quite large and may be arranged as subclusters that facilitate jumps of propagating local Ca^{2+} fronts over sub-micron distances [19]. Cluster diameters estimated therein in the presence of the calcium chelator EGTA were ~400–900 nm, and the side-to-side spacing between adjacent sites ~500 nm. Other experimental studies have also pointed toward a similar picture in mammalian cells [16]. According to our estimations, propagation of Ca^{2+} from the release site to neighbor clusters during a typical event taking place in the presence of the fluorescent indicator can lead to an increase of ~150–300 nM free Ca^{2+} within ~30 ms at the level of an adjacent release site situated at 0.5–1 μm away from the original releasing site. The corresponding Ca^{2+} level that would be established in the absence of the calcium indicator is ~250–500 nM, which resides within the activating range of the IP_3 receptor even at low concentrations of IP_3 [8, 33]. However, in their experiments Sun et al. [5] used medium IP_3 levels in order to prevent the formation of calcium waves and hence observe delimited individual puffs. Therefore it is likely that the
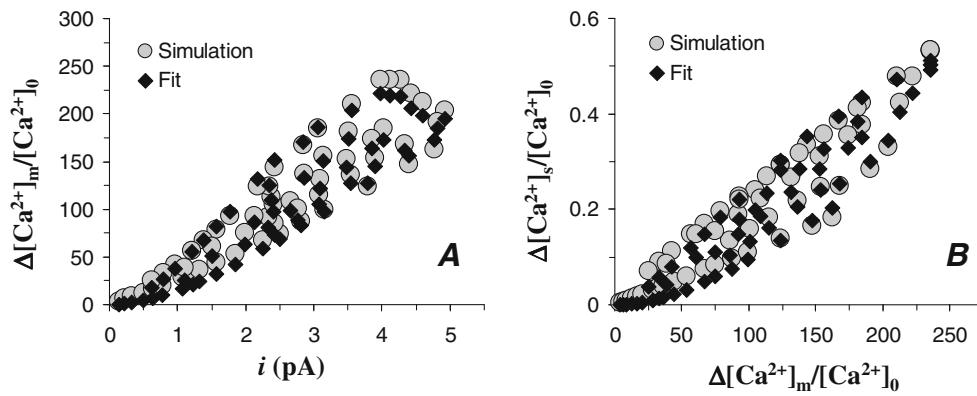
Fig. 8 Increase in maximal Ca^{2+} levels near the release site ($\Delta[Ca^{2+}]_m$) or in a 783.74 fl domain surrounding the release site ($\Delta[Ca^{2+}]_s$) obtained from the same numerical simulations as in Fig. 3. In (**a**) the relation between the increase $\Delta[Ca^{2+}]_m$ relative to the resting level [Ca^{2}

$^{+}]_0$ and the release current i is shown. In (**b**) the relation between the relative increases $\Delta[Ca^{2+}]_m/[Ca^{2+}]_0$ and $\Delta[Ca^{2+}]_s/[Ca^{2+}]_0$ is depicted. Simulation data (circles) are displayed together with the fitting results (diamonds) obtained from Eqs. 26–27

actual calcium events evolve normally at lower release rates than in the presence of the indicator which can reduce the concentration of Ca^{2+} at the channel mouth and thus increase the local Ca^{2+} gradient across the ER membrane [28]. Consequently, the intrinsic activity duration of a given IP$_3$R cluster is likely to decrease when an exogenous buffer is added [28] because inactivating Ca^{2+} levels at the channel mouth can be reached faster. This idea is strongly supported by various experimental data obtained with 40 μM Oregon Green-1 in the absence or the presence of 300 μM EGTA [5,

Fig. 9 Predicted distributions of (**a**) maximal Ca^{2+} levels near the release site ([Ca^{2+}]$_m$) or (**b**) average Ca^{2+} levels in the 783.74 fl domain surrounding the release site ([Ca^{2+}]$_s$) relative to the resting level [Ca^{2+}]$_0$, corresponding to the animal ("AH") or the vegetal hemisphere ("VH") of the oocyte

16, 18, 19], which indicate a strong reduction of the release duration induced by EGTA (from ~100 ms to ~20 ms). In addition, other recent experimental and modeling studies indicate a release duration of ~60 ms obtained with 40 μM Oregon Green-1 in the presence of a lower level of 135 μM EGTA [29], and about 18 ms obtained with 300 μM EGTA and 25 μM Fluo-4 Dextran, which is a low affinity calcium indicator [18, 25, 29].

In agreement with earlier experimental findings, there appears to be a consistent difference between the two hemispheres of the oocyte regarding the distribution of IP$_3$ receptors and the magnitude of calcium release events. The variations in the distributions of liberated Ca^{2+} between the animal and the vegetal hemispheres (Fig. 1) are probably due to the differing number of active receptors within a releasing cluster, or alternatively the IP$_3$R clusters may comprise different isoforms of IP$_3$ receptors. Such functional differences between Ca^{2+} release sites in the two hemispheres appear to play a critical role both at fertilization and in the early development phases of the *Xenopus* embryo. In immature *Xenopus* oocytes, the peripheral endoplasmic reticulum forms an intricate network of interconnected cisternae, linear tubules, polygonal reticulum and three-way junctions generated by fusion of intersecting tubules, which appears to be restricted to a ~6–8 μm-wide shell situated a few micrometers under the plasma membrane [15, 20]. The observed tubular/cisternal structures of the peripheral ER in the large sized oocyte have variable diameters, from narrow tubules (50–100 nm width) to tubes, cisternae or junctions that appear to be several times wider [34–37]. In addition, the subcortical region of the oocyte observed at high resolution in electron micrographs has been found to contain long, dense stacks of cisternae (the annulate lamellae) of ~2–4 μm width and ~20 μm in length [34, 37], from which ER cisternae migrate to the cortex during the maturation process.

Moreover, a few micrometers underneath the plasma membrane, there are also large cortical granules of ~1–2 μm diameter and smaller pigment and glycogen granules of ~0.5 μm width, dispersed at relatively small distances (~100 nm) one from another [15, 34, 35]. During maturation, membrane junctions form between the plasma membrane and cortical endoplasmic reticulum, which represent sites for the transduction of extracellular events into intracellular calcium release during fertilization and activation of development [38]. Such junctions exhibit a distinctive animal-vegetal polarity of distribution [38] which is correlated with the hemispheric asymmetry of the calcium-release capacity of the ER calcium stores. After fertilization, a calcium wave propagates inside the egg [39, 40] which in turn triggers a wave of cortical granule exocytosis (with the primary role in creating the fertilization envelope which provides a long-term block to polyspermy) and an extensive contraction of the cortex which propagates faster in the animal hemisphere as compared with the vegetal hemisphere [34, 38]. A higher concentration of free calcium in the animal hemisphere could account for the latter phenomenon [38], as well as for the higher rate of intracellular Ca^{2+} increase during propagation of the calcium wave in the animal versus the vegetal half of the egg [39]. Likewise, the rapid mitoses in early *Xenopus* embryos are accompanied by two surface contraction waves which travel from the animal to the vegetal pole ahead of the developing cleavage furrow [41] and are in close spatiotemporal correlation with the two waves of the mitosis-promoting factor (MPF) activation and inactivation

initiated by components in the animal cytoplasm. Since metaphase to anaphase transition is a calcium-dependent process [40, 42, 43], the involvement of Ca^{2+} release events in the mitotic program may be associated with the functional role of the observed asymmetry between the two hemispheres of the oocyte.

Peak intracellular Ca^{2+} levels reached at deep subcortical regions during the propagation of the calcium wave elicited after fertilization are ≈ 1.2 μM [39]. However, it appears that cortical granule exocytosis requires free calcium levels of the order of 2–5 μM [39], whereas activation of plasma membrane Ca^{2+}-activated Cl^- channels requires >27 μM free Ca^{2+} for half-maximal activation [44]. Activation of the Cl^- channels after fertilization provides a fast block to polyspermy via depolarization of the egg (the fertilization potential) which prevents further sperm fusion with the egg membrane [40]. Consistent with the polarized distribution of the Ca^{2+} release sites, a clear hemispheric asymmetry also manifests with regard to the amplitudes or kinetics of calcium-activated Cl^- currents [14, 45]. Our numerical studies indicate that in the absence of an exogenous calcium indicator, average Ca^{2+} levels elicited following medium-intensity stimulation with IP_3 can reach ~2 μM and ~0.4 μM at ~300 nm from a given release site situated in the animal or the vegetal hemisphere, respectively, while in less frequent events the corresponding maximal Ca^{2+} level can reach values as high as 11 μM and 4 μM, respectively. Hence, these figures appear to be realistic and consistent with the idea that after the full IP_3R activation that accompany the development of the calcium wave inside the fertilized egg, sufficiently high levels of free Ca^{2+} can build up close to the cortical ER release sites and thus can trigger the exocytosis of the cortical granules or the activation of plasma membrane Cl^- currents.

References

1. Clapham DE (1995) Calcium signalling. Cell 80:259–268
2. Berridge MJ, Bootman MD, Lipp P (1998) Calcium - a life and death signal. Nature 395:645–648
3. Case RM, Eisner D, Gurney A, Jones O, Muallem S, Verkhratsky A (2007) Evolution of calcium homeostasis: from birth of the first cell to an omnipresent signalling system. Cell Calcium 42:345–350
4. Mak DO, Foskett JK (1997) Single-channel kinetics, inactivation, and spatial distribution of inositol (1,4,5) trisphosphate (IP_3) receptor in Xenopus oocyte nucleus. J Gen Physiol 109:571–587
5. Sun XP, Callamaras N, Marchant JS, Parker I (1998) A continuum of $InsP_3$-mediated elementary Ca^{2+} signalling events in Xenopus oocytes. J Physiol 509:67–80
6. Thomas D, Lipp P, Berridge MJ, Bootman MD (1998) Hormone stimulated elementary Ca^{2+} signals are not stereotypic, but reflect activation of different size channel clusters and variable recruitment of channels within a cluster. J Biol Chem 273:27130–27136
7. Falcke M (2004) Reading the patterns in living cells - the physics of Ca^{2+} signaling. Adv Phys 53:255–440
8. Foskett JK, White C, Cheung KH, Mak DOD (2007) Inositol trisphosphate receptor Ca^{2+} release channels. Physiol Rev 87:593–658
9. Schuster S, Marhl M, Höfer T (2002) Modelling of simple and complex calcium oscillations. From single-cell responses to intercellular signalling. Eur J Biochem 269:1333–1355
10. Skupin A, Kettenmann H, Falcke M (2010) Calcium signals driven by single channel noise. PLoS Comput Biol 6:e1000870
11. Perc M, Rupnik M, Gosak M, Marhl M (2009) Prevalence of stochasticity in experimentally observed responses of pancreatic acinar cells to acetylcholine. Chaos 19:037113
12. Baran I (2003) Integrated luminal and cytosolic aspects of the calcium release control. Biophys J 84:1470–1485
13. Perc M, Green AK, Dixon CJ, Marhl M (2008) Establishing the stochastic nature of intracellular calcium oscillations from experimental data. Biophys Chem 132:33–38
14. Callamaras N, Sun XP, Ivorra I, Parker I (1998) Hemispheric asymmetry of macroscopic and elementary calcium signals mediated by $InsP_3$ in *Xenopus* oocytes. J Physiol Lond 511:395–405
15. Callamaras N, Parker I (1999) Radial localization of inositol 1,4,5-trisphosphate-sensitive Ca^{2+} release sites in *Xenopus* oocytes resolved by axial confocal linescan imaging. J Gen Physiol 113:199–213
16. Smith IF, Parker I (2009) Imaging the quantal substructure of single IP_3R channel activity during Ca^{2+} puffs in intact mammalian cells. Proc Natl Acad Sci USA 106:6404–6409
17. Taufiq-Ur-Rahman SA, Falcke M, Taylor CW (2009) Clustering of $InsP_3$ receptors by $InsP_3$ retunes their regulation by $InsP_3$ and Ca^{2+}. Nature 458:655–659
18. Rose HJ, Dargan S, Shuai J, Parker I (2006) 'Trigger' events precede calcium puffs in *Xenopus* oocytes. Biophys J 91:4024–4032
19. Demuro A, Parker I (2008) Multi-dimensional resolution of elementary Ca^{2+} signals by simultaneous multi-focal imaging. Cell Calcium 43:367–374
20. Marchant JS, Ramos V, Parker I (2002) Structural and functional relationships between Ca^{2+} puffs and mitochondria in *Xenopus* oocytes. Am J Physiol Cell Physiol 282:C1374–C1386
21. Zou H, Lifshitz LM, Tuft RA, Fogarty KE, Singer JJ (2004) Using total fluorescence increase (signal mass) to determine the Ca^{2+} current underlying localized Ca^{2+} events. J Gen Physiol 124:259–272
22. Shuai J, Parker I (2005) Optical single-channel recording by imaging Ca^{2+} flux through individual ion channels: theoretical considerations and limits to resolution. Cell Calcium 37:283–299
23. Izu LT, Wier WG, Balke CW (1998) Theoretical analysis of the Ca^{2+} spark amplitude distribution. Biophys J 75:1144–1162
24. Smith GD, Keizer JE, Stern MD, Lederer WJ, Cheng H (1998) A simple numerical model of calcium spark formation and detection in cardiac myocytes. Biophys J 75:15–32
25. Shuai J, Rose HJ, Parker I (2006) The number and spatial distribution of IP_3 receptors underlying calcium puffs in *Xenopus* oocytes. Biophys J 91:4033–4044
26. Baran I (2007) Characterization of local calcium signals in tubular networks of endoplasmic reticulum. Cell Calcium 42:245–260

27. Baran I (2008) Modulation of calcium signals by fluorescent dyes in the presence of tubular endoplasmic reticulum: A modelling approach. BioSystems 92:259–269

28. Baran I, Popescu A (2009) A model-based method for estimating Ca^{2+} release fluxes from linescan images in *Xenopus* oocytes. Chaos 19:037106

29. Bruno L, Solovey G, Ventura AC, Dargan S, Dawson SP (2010) Quantifying calcium fluxes underlying calcium puffs in *Xenopus laevis* oocytes. Cell Calcium 47:273–286

30. Vais H, Foskett JK, Mak DOD (2010) J Gen Physiol 136:687–700

31. Hille B (1992) Ionic Channels of Excitable Membranes. Sunderland, MA, Sinauer Associates

32. Baran I, Iftime A, Popescu A (2010) Diffusion - convection effects on drug distribution at the cell membrane level in a patch-clamp setup. BioSystems 102:134–147

33. Baran I (2005) Gating mechanisms of the type-1 inositol 1,4,5-trisphosphate receptor. Biophys J 89:979–998

34. Campanella C, Andreucetti P, Taddei C, Talevi R (1984) The modifications of cortical endoplasmic reticulum during in vitro maturation of *Xenopus laevis* oocytes and its involvement in cortical granule exocytosis. J Exp Zool 229:283–293

35. Jafri MS, Vajda S, Pasik P, Gillo B (1992) A membrane model for cytosolic calcium oscillations. A study using Xenopus oocytes. Biophys J 63:235–246

36. Voeltz GK, Rolls MM, Rapoport TA (2002) Structural organization of the endoplasmic reticulum. EMBO Rep 3:944–950

37. Terasaki M, Runft LL, Hand AR (2001) Changes in organization of the endoplasmic reticulum during *Xenopus oocyte* maturation and activation. Mol Biol Cell 12:1103–1116

38. Gardiner DM, Grey RD (1983) Membrane junctions in *Xenopus* eggs: their distribution suggests a role in calcium regulation. J Cell Biol 96:1159–1163

39. Busa WB, Nuccitelli R (1985) An elevated free cytosolic Ca^{2+} wave follows fertilization in eggs of the frog, *Xenopus laevis*. J Cell Biol 100:1325–1329

40. Machaca K (2007) Ca^{2+} signaling differentiation during oocyte maturation. J Cell Physiol 213:331–340

41. Pérez-Mongiovi D, Chang P, Houliston E (1998) A propagated wave of MPF activation accompanies surface contraction waves at first mitosis in *Xenopus*. J Cell Sci 111:385–393

42. Baran I (1996) Calcium and cell cycle progression: possible effects of external perturbations on cell proliferation. Biophys J 70:1198–1213

43. Reber S, Over S, Kronja I, Gruss OJ (2008) CaM kinase II initiates meiotic spindle depolymerization independently of APC/C activation. J Cell Biol 183:1007–1017

44. Gomez-Hernandez JM, Stuhmer W, Parekh AB (1997) Calcium dependence and distribution of calcium-activated chloride channels in *Xenopus* oocytes. J Physiol 502:569–574

45. Lupu-Meiri M, Shapira H, Oron Y (1988) Hemispheric asymmetry of rapid chloride responses to inositol trisphosphate and calcium in *Xenopus* oocytes. FEBS Lett 240:83–87

ORIGINAL PAPER

Chemisorptions effect of oxygen on the geometries, electronic and magnetic properties of small size Ni_n (n = 1-6) clusters

Debashis Bandyopadhyay

Abstract The present study reports the effect of oxygen addition on small size Ni_n (n=1-6) clusters in different spin states within the framework of linear combination of atomic orbital (LCAO) density functional theory (DFT) under spin polarized generalized gradient approximation (GGA) functional. Relative stabilities of the optimized clusters are discussed on the basis of the calculated parameters, such as, binding energy (BE), embedding energy (EE) and fragmentation energy (FE). Other parameters, like ionization potential (IP), electron affinity (EA), etc. show that though the additions of oxygen can affect the chemical properties of Ni_n clusters with an additional stability to Ni_nO. In most of the cases the magnetic moment of the stable isomers are geometry dependent for a particular size both in pure and oxidized clusters. Calculated magnetic moments of Ni_nO (n=1-6) clusters reveal that the magnetic moment of ground state Ni_4O isomers in different geometries is same as in pure Ni_4 isomers. Present study also explains the cause of stable magnetic moment in Ni_4O cluster through the distribution of electrons in different orbitals.

Keywords Chemical properties · Classification codes:20.040 · 30.180 · 40.080 · Clusters · Density functional theory · Nanomaterials

D. Bandyopadhyay (✉)
Physics Department, Birla Institute of Technology and Science,
Pilani 333031 Rajasthan, India
e-mail: Debashis.bandy@gmail.com

D. Bandyopadhyay
e-mail: bandy@bits-pilani.ac.in

Introduction

Since the last few decades transition metal (TM) oxide clusters are becoming interesting field of theoretical as well as experimental research because of their novel electronic, magnetic and optical properties as well as the prominent role in the catalytic activities, environmental process in addition to several other applications in nanoscience and nanotechnology [1–12]. Although there are many experimental studies by using mass spectrometry and photoelectron spectroscopy of TM oxides, but determining the relative stabilities of these systems remains problematic. Several experimental works on TM oxide cluster in the gas phase have contributed fundamental information about their chemical bonding, reactivity, magnetic, and electronic properties. Herman et al. [13] studied the gas phase reaction of different transition metal ions (Ti^+, V^+, Fe^+, Co^+, Ni^+, Cu^+, Zn^+) with CO and CO_2 followed by *ab initio* calculations for further understanding of reaction mechanism, reaction kinetics and thermochemistry of the nanoclusters. Xu et al. [14] studied the oxygen affinity of pure Pt_n clusters (n=1-10) using density functional theory (DFT). By using laser vaporization and electro-spray experimental techniques followed by photoelectron spectroscopy Zhai et al. [15] studied electronic and geometric structures of doubly and singly charged species $M_2O_7^{2-}$, $MM'O_7^{2-}$, and $M_2O_7^-$ (M, M' = Cr, Mo, W) followed by DFT calculations to understand the experimental evolution of geometric and electronic structures as a function of charge state. Liu et al. [16] studied the effect of addition of oxygen on the small size Co clusters. They found that the addition of oxygen could not produce much effect on the average magnetic moment of the Co clusters. Li et al. [17] calculated total atomization energies and normalized clustering energies of $(MO_2)_n$ (M = Ti, Zr, Hf) and $(MO_3)_n$

(M = Cr, Mo, W) transition metal oxide clusters up to $n=4$ at the coupled cluster [CCSD(T)] and density functional theory (DFT) levels. In another study Zhai et al. [18] reported a comparative study of reduced transition metal oxide clusters, $M_3O_8^-$ (M = Cr, W) in neutral and anionic state via anion photoelectron spectroscopy, density functional theory, molecular orbital theory (CCSD(T)) calculations and predicted the redox reaction in thermochemistry of the system.

Among several other transition metal oxides, small size Ni_nO clusters is one of the most important candidates because of its high magnetic moment and stability. Nickel oxide is a highly insoluble thermally stable nickel source suitable for glass, optics and ceramic applications. Oxide compounds are usually not electrical conductor; however certain nickel oxides are electronically conductive and therefore have useful applications in fuel cells and oxygen generation systems where they exhibit ionic conductivity. Moreover, nickel oxide is also an important element as a component in lightweight aerospace applications. Several other theoretical [19–24] and experimental [25–30] studies on nickel oxides in neutral and charged states are also reported. Most of the theoretical studies are focused on the geometrical, electronic and magnetic properties of pure and oxidized nickel clusters Ni_n ($n \leq 60$) [19–23]. On the other hand, most of the experimental studies [25–30] on nickel oxide clusters are focused on their magnetic and chemical properties by using gas phase reaction, photoelectron spectroscopy and Mössbauer spectroscopy. Both theoretical and experimental research on Ni_nO clusters are challenging because of its complicated electronic and magnetic properties due to the 3d sub-shells and in addition, the energy difference between different spin states is very small. Therefore, determination of the ground state in a particular size is a difficult work. In this report a detailed theoretical investigation of small size Ni_n and Ni_nO ($n=1-6$) clusters of different geometries and spin states (singlet to 11) is presented to understand their electronic as well as magnetic behavior and their variation with the geometry, size and spin states in a systematic manner. In the latter part, both Ni_n and Ni_nO ($n=1-6$) optimized ground state geometry in each size are then selected to study their binding energy, embedding energy, stability, chemical potential, ionization potential etc. Comparisons of the present calculations with the available experimental and theoretical results are also done.

Computational

In the present report geometry optimization of Ni_n and Ni_nO ($n=1-6$) clusters are performed by using density-functional theory (DFT) with the unrestricted B3PW91 exchange-correlation potential [31–34]. Self-consistent field (SCF) electronic structure calculations were carried out on all clusters within the framework of Kohn-Sham DFT [35]. Molecular orbital (MO) are expressed as linear combination of atom-centered basis functions for which the standard Gaussian LanL2DZ basis set and associated effective core potential (ECP) is used on all the atoms. Spin-polarized calculations are carried out using the Becke three-parameter exchange and the Perdew-Wang generalized gradient approximation (GGA) functional [36–39]. The standard LanL2DZ basis sets are employed to provide an effective way to reduce difficulties in calculations of two-electron integrals caused by transition metal Ni atom [40]. In order to obtain the lowest-energy Ni_nO structures, we have chosen a considerable number of possible structures as initial geometries in each cluster size ($n=1-6$). Based upon the available theoretically and experimentally verified geometries of Ni_n [23] different evolution patterns for determining the Ni_nO clusters, including Ni-substituted and O-capped patterns, are first taken into account and then the 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 stability is reassured by calculating the frequency of harmonic vibration. If any imaginary frequency is found, a relaxation along that vibrational mode is carried out until the true local minimum is obtained. In all clusters, geometries were optimized with no symmetry constraints in each initial guess structure. Both in case of pure and oxygenated nickel clusters optimizations are done with the spin multiplicity varied from singlet to 11 and are independent of the size of the clusters. Optimized ground state geometries in each size with their spin multiplicity are shown in Fig. 1. In case, when the total optimization energy decreases with increasing spin, increasingly higher spin states is considered until the energy minimum with respect to spin is reached. To check the reliability of the present level of calculation, calculated parameters of Ni_2 and NiO dimers, as example, the bond length, ionization potential, electron affinity and the lowest vibrational frequency are compared with the reported experimental and theoretical data available as shown in Tables 1 and 2. It shows that the present calculated values are comparable to the reported values and hence it is expected that the present level of calculation can be applied to a bigger size clusters in the same system. All theoretical calculations are carried out with the Gaussian 03 program package [41].

Results and discussion

Optimized structures of different bare nickel and nickel oxide clusters calculated in the present study are shown in

a

b
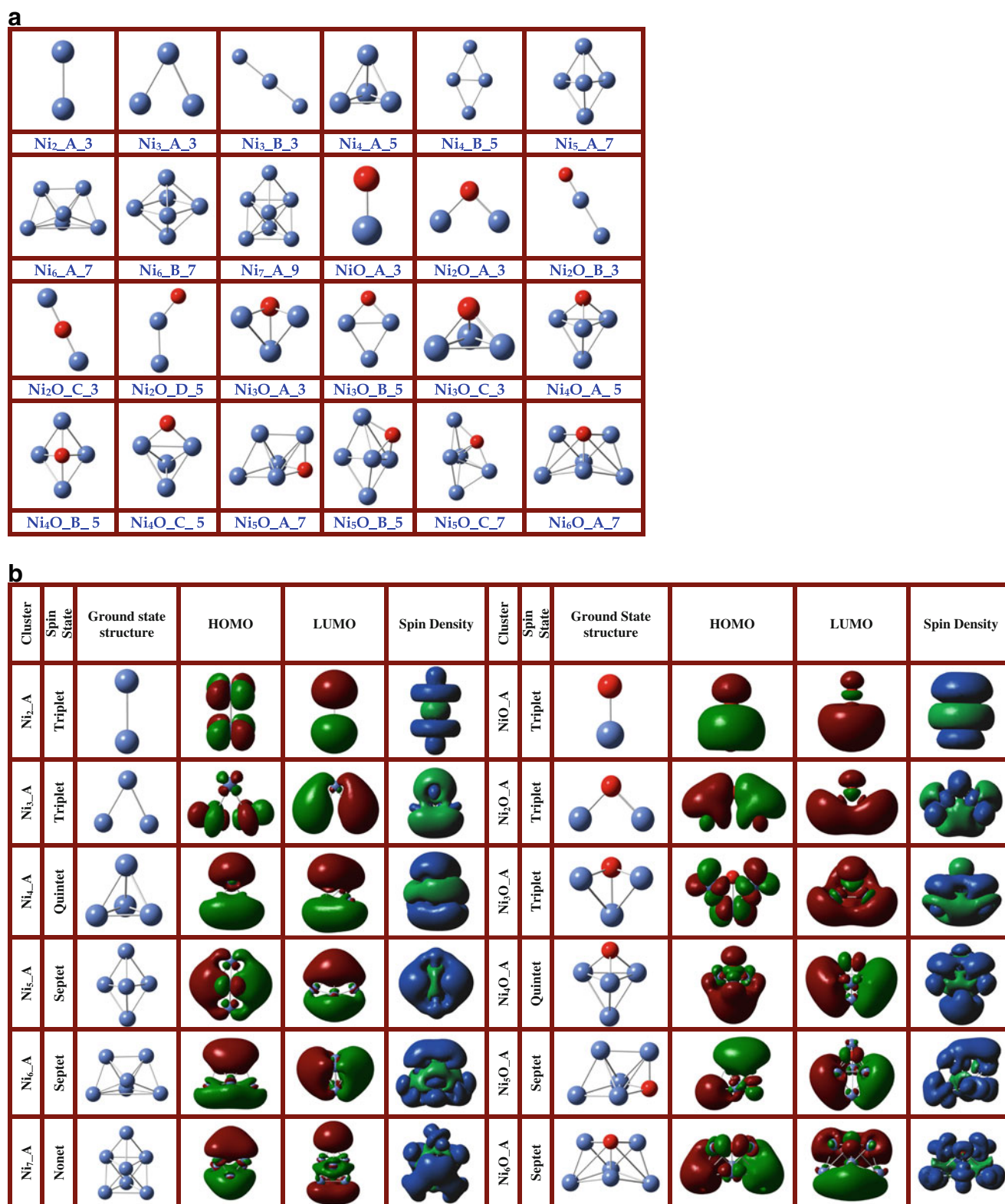
Fig. 1 (a) Optimized structures of Ni_n and Ni_nO clusters. In the figure, the last digit in name represents the spin multiplicity of the clusters, as: Triplet (3), Quintet (5), Septet (7) and Nonlet (9). All clusters shown in figure, correspond to the minimal energy among the different structural isomers, and the different spin multiplicities studied in a particular size. (b) Orbital and spin density distribution of ground state Ni_n and Ni_nO clusters among different structural isomers and the different spin multiplicities (singlet to nonet) studied of different sizes. All clusters shown in figure, correspond to the minimal energy among the different structural isomers, and the different spin multiplicities studied in a particular size

Table 1 Different parameters of pure and charged Ni₂ and NiO clusters

Dimers	Bond length (Å)	Lowest frequency (cm⁻¹)	IP (eV)	EA (eV)
Ni2	2.06[23], 2.155[42], 2.13[21], 2.20[45][‡], 2.36[a]	210±25[42], 236[a]	7.14[46], 7.06[a]	0.9262±0.01[42] 1.00[a]
NiO	1.626[47], 1.627[48], 1.63[44][‡], 1.89[49], 1.70[a]	800[43], 839[44], 836[a]		1.46[43], 1.50[a]
Ni2-	2.257±0.017[42] 2.37[a]	280±20[42], 236[a]		
NiO-	1.68[a]	810[43], 784[a]		

[‡] Experimental, [a] Present work: B3PW91/LanL2DZ

Fig. 1a-b along with their HOMO-LUMO orbital distributions and total spin densities. Different physical and chemical parameters of the pure and doped clusters such as bond length, frequency, ionization potential, point group symmetry, electronic states, relative energies etc. are shown in Tables 1 2 and 3. In the following section geometries, stabilities, electronic and magnetic properties of the nanoclusters will be discussed on the basis of the calculated parameters and their variation with the cluster size.

Geometries and stabilities

Ni₂ and NiO clusters

As the first members in the series, different physical and chemical parameters of Ni₂ (D$_{\alpha h}$) and NiO dimers are calculated and presented in Tables 1 and 2 with the reported theoretical and experimental data. Variation of the optimized energies of NiO dimer in different spin states are presented in Table 3 with respect to the optimized ground state dimer with C$_{\infty v}$ point group symmetry and $^3\Sigma^+$ electronic state. In the present calculation the bond length of the pure nickel dimer varies with different spin multiplicities from singlet to 11 in the range of 2.32Å to 2.45Å with a value of 2.36Å in triplet ground state, which is comparable to the experimental values presented in Table 1. Similarly, the bond length of NiO cluster is found as 1.70Å. This is close to the experimental value 1.63Å.

Table 2 IP (eV) from reported and present calculations

Reference	Ni₂	Ni₃	Ni₄	Ni₅	Ni₆
[23]	8.3	6.8	5.9	6.6	6.8
[50]	5.4	3.9	4.4	4.7	4.9
[51]			6.08	5.83	6.65
[52]	5.7	4.2	5.1	4.8	4.4
[53]	6.5	6.2	6.2	5.9	5.6
[54]	6.8	6.5	6.3	6.3	6.3
Ref. [51, 52] in ref [23]	7.6	6.12	5.70	6.20	6.78
Present	7.06	5.55	5.46	6.31	6.77

Other calculated parameters, as like, IP and EA are in good agreement with the reported theoretical and experimental data shown in Tables 1 and 2. From the parameters presented in Tables 1 and 2, it is clear that present level of calculation is appropriate for both pure nickel and nickel oxide clusters. In both the clusters calculated natural vibrational frequencies are also in good agreement with the previously reported values. The ionization potential calculated for Ni₂ cluster is 7.06 eV and this is within the range of reported experimental and theoretical values that varies from 5.4 eV to 8.3 eV as shown in Table 2. Electron affinities of Ni₂ and NiO clusters are also quite close to the reported values shown in Table 1. Calculated HOMO and LUMO orbital distributions of the optimized ground state of Ni$_n$ (n=1-7) and Ni$_n$O (n=1-6) clusters along with the total spin magnetic moment distributions are shown in Fig. 1b. In Ni₂ dimer structure, the HOMO distribution is like a double quadruple charge distribution (π-bonding) whereas the LUMO orbital is like a perfect dumbbell structure (σ-bonding). In NiO dimer, most of the orbital charge density distributed around the nickel atom and this is even more in LUMO orbital. The total spin density distribution is symmetrical in Ni dimer but it is more around the nickel atom in NiO cluster as shown in Fig. 1b.

Ni₃ and Ni₂O cluster

In bare Ni₃ cluster two different geometries are obtained. The ground state geometry is a triangle with internal angle close to 56.7° in triplet spin state with C$_s$ point group symmetry. This structure is lower in energy than the linear chain structure by an amount 0.06 eV. In linear chain geometry the ground state is a quintet with D$_{\infty h}$ point group symmetry. At this size the spin state in different ground state geometries is geometry dependent. The same behavior is also found in other systems and will be discussed in afterward section. By replacing the end and the middle Ni atoms by oxygen in the ground state Ni₃_A and in Ni₃_B (Fig. 1a), four different optimized structures at different spin states are obtained. The first triangular geometry (Ni₂O_A) is the ground state structure in triplet spin state. Out of the other three remaining structures, two are linear chain like (Ni-O-Ni or Ni-Ni-O) and the third one is a bend

structure (C_S symmetry) where the Ni-Ni-O angle is 109^0 as shown in Fig. 1a. The first linear chain structure Ni-Ni-O has a $C_{\infty v}$ point group symmetry, whereas the 2nd linear chain structure with an oxygen at the middle position of the chain (Ni-O-Ni) has $D_{\infty h}$ symmetry. Here the first and third structures of Ni_2O are in triplet state, whereas, the second and fourth structures are in quintet and singlet states respectively. It is to be noted that the ionized Ni_2O^+ cluster is a linear chain structure with the oxygen atom at the middle (Ni-O-Ni). In Ni_2O^+ the Ni atoms contribute charge in ionization. Calculated ionization potential of Ni_3 cluster is in good agreement with the reported value. The reported value of ionization potential varies from 3.9 eV to 6.8 eV (Table 2) whereas the present value is 5.54 eV.

Ni_4 and Ni_3O clusters

Between two different isomers of Ni_4, the triangular based pyramid geometry is the optimized ground state in quintet spin state with D_{2d} point group symmetry and $^5A_2$ molecular orbital electronic state represented as irreducible representation (IR) in Table 3. The other structure is a bend rhombus. The ionization potential of the ground state structure is 5.46 eV is close to reported values as given in Table 2. Replacing one-nickel atom in bend rhombus isomer of Ni_4 structure, Ni_3O optimized structure with C_{3v} point group symmetry is obtained as shown in Fig. 1a. In the present calculation the linear chain structures are not considered because of the presence of an imaginary frequency. Relative energies of the structures in different spin state with respect to the ground state structure are presented in Table 3.

Ni_5 and Ni_4O clusters

Two different optimized structures are obtained in Ni_5 size. The distorted Ni capped rhombus based structure is the ground state geometry with C_s point group symmetry in septet spin state. The other structure is a bi-capped triangular Ni_3 structure (not shown). A number of stable geometries are obtained in Ni_4O cluster as shown in Fig. 1a. The ground state geometry is obtained by replacing a capped nickel atom from the bi-capped Ni_3 triangular structure by oxygen. The ground state of Ni_4O is in a quintet spin state as like ground state Ni_4 isomers. The binding energy of this structure is higher than Ni_4 or Ni_5 structures.

Ni_7, Ni_6, Ni_5O and Ni_6O clusters

The most stable structure in Ni_6 is bi-capped quadrilateral with C_s point group symmetry. By replacing one nickel with oxygen, three different kinds of optimized structures are obtained. Among them, bi-capped Ni rhombus structure in septet spin state is the ground state geometry with C_s point group symmetry (Fig. 1b). The ground state geometry of Ni_5O structure is in septet. Three other ground state geometries are also obtained in Ni_5O series in triplet; quintet and septet spin states respectively. Energetically these three structures are very close to the overall ground state structure within the energy difference of 0.05 eV. Addition of a Ni or an oxygen atom on the surface of Ni_6 pure cluster gives optimized slightly distorted Ni_7 and Ni_6O structures respectively. Optimized Ni_7 cluster is a nonet and Ni_6O is in septet spin state. Different orbital (HOMO and LUMO) and spin distributions of Ni_n and Ni_nO ground state structures are shown in Fig. 1b.

It is clear from the above observation that for the clusters from n=2 to 6, except for n=4, the spin state is geometry dependent in a particular size as shown in Table 3. It is commonly observed that in these geometries of different sizes, the structures where O atom is bonded with three-nickel atoms are the most stable. In these structures oxygen atom prefers to sit as surface capped atom on the Ni_n clusters.

The relative stabilities, electronic and magnetic properties

In order to get the idea about the relative stabilities, electronic and magnetic properties of Ni_nO (n=1-6) clusters, different parameters like, binding energy (BE), HOMO-LUMO gap, embedding energy (EE), fragmentation energy (FE or $\Delta(n,n-1)$), stability or second order change in energy ($\Delta_2(n)$), ionization potential (IP), etc. are calculated. Variation of these parameters with the size of Ni_nO and Ni_n are then plotted and compared. On the basis of these calculated parameters and their variations with the size, stabilities and other properties of the clusters are discussed.

In the present work the following relation defines binding energy per atom of a cluster:

$$BE = \frac{[E(Ni_nO) - E(O) - nE(Ni)]}{n+1} \, or \quad (1)$$
$$\frac{[E(Ni_n) - nE(Ni)]}{n}$$

for Ni_nO and Ni_n respectively. Here, $E(Ni_nO)$ and $E(Ni_n)$ is the energy of the Ni_nO and Ni_n clusters respectively, $E(O)$ is the energy of the isolated oxygen atom; $E(Ni)$ is the energy of a nickel atom. Variation of binding energy curve is shown in Fig. 2 for pure and oxidized nickel clusters. It is clear from Fig. 2 that with the increase of n during the growth process of the cluster, the binding energy increases

Table 3 Optimized parameters of Ni_nO (n=1-6) isomeric clusters in different spin states

Cluster	Isomer	Multiplicity	Relative energy (eV)	Symmetry	IR[*]	DM[**] (debye)	HOMO-LUMO gap (eV)
NiO	A	1	1.61	$C_{\infty v}$	$^1\Sigma^+$	5.67	1.33
		3	0.00		$^3\Sigma^+$	5.24	3.52
		5	1.28		$^5\Sigma^+$	0.24	4.59
		7	6.57		$^7\Sigma^+$	1.78	1.3
		9	9.68		$^9\Sigma^+$	5.87	1.78
		11	11.95		$^{11}\Sigma^+$	15.03	4.08
Ni_2O	A	1	0.76	C_{2v}	$^1A_1$	3.22	2.24
		3	0.00		$^3A_2$	3.25	3.6
		5	0.43		$^5A_2$	0.13	1.47
		7	2.27	$C_{\infty v}$	$^7\Sigma^+$	1.25	3.2
		9	6.76	C_s	$^9A'$	0.11	1.06
		11	12.1		$^{11}A'$	2.38	1.36
Ni_2O	B	1	2.73	$C_{\infty v}$	$^1\Sigma^+$	6.45	1.84
		3	0.18		$^3\Sigma^+$	4.86	4.13
		5	0		$^5\Sigma^+$	5.47	4.93
		7	2.63		$^7\Sigma^+$	0.21	2.36
		9	6.47		$^9\Sigma^+$	0.23	1.23
		11	11.28		$^{11}\Sigma^+$	2.36	1.72
Ni_2O	C	1	1.04	$D_{\infty h}$	$^1\Sigma_u$	0.00	1.74
		3	0.00		$^3\Sigma_u$	0.00	3.67
		5	0.91		$^5\Sigma_u$	0.00	1.36
		7	1.75		$^7\Sigma_u$	0.00	4.07
		9	8.00		$^9\Sigma_u$	0.00	0.99
		11	12.7	C_{2v}	$^{11}A_2$	0.00	0.7
Ni_2O	D	1	0	C_s	$^1A_1$	3.23	2.23
		3	0.23		$^3A''$	4.27	4.04
		5	0.09		$^5A''$	4.75	4.41
		7	2.31		$^7A''$	2.47	2.6
		9	6.24		$^9A'$	0.17	0.84
		11	11.34		$^{11}A'$	2.58	1.38
Ni_3O	A	1	1.87	C_s	$^1A'$	2.37	1.54
		3	0.00	C_{3v}	$^3A_2$	2.67	2.95
		5	0.35		$^5A_1$	2.02	2.1
		7	1.39		$^7A_2$	1.21	1.54
		9	3.35		$^9A_2$	0.01	3.33
		11	7.18		$^{11}A_2$	0.62	1.15
Ni_3O	B	1	1.45	C_s	$^1A'$	3.13	2.18
		3	0		$^3A''$	4.11	2.57
		5	0.1		$^5A''$	4.52	2.86
		7	0.6		$^7A''$	2.48	2.03
		9	3.41		$^9A'$	1.71	2.52
		11	6.86		$^{11}A''$	0.42	1.61
Ni_3O	C	1	1.87	C_s	$^1A'$	2.38	1.54
		3	0.00	C_3	^3A	2.67	2.95
		5	0.38		^5A	2.22	2.12
		7	1.27	C_s	$^7A'$	1.19	1.59
		9	2.71		$^9A''$	1.41	3.25
		11	7.19	C_{3v}	$^{11}A_1$	0.71	1.19
Ni_3O	D	1	0.00	C_{2v}	$^1A'$	3.23	2.16

Table 3 (continued)

Cluster	Isomer	Multiplicity	Relative energy (eV)	Symmetry	IR[*]	DM[**] (debye)	HOMO-LUMO gap (eV)
		3	1.09	C_s	$^3A''$	4.27	2.01
		5	0.96		$^5A''$	4.74	2.04
		7	1.62		$^7A'$	2.47	3.87
		9	1.75		$^9A'$	0.17	2.45
		11	7.45		$^{11}A''$	2.58	0.6
Ni_4O	A	1	2.48	C_3	^1A	2.56	1.22
		3	0.01	C_{3v}	$^3A_2$	3.39	2.47
		5	0.00	C_s	$^5A'$	3.29	2.51
		7	0.87	C_1	^7A	1.98	1.54
		9	1.71		^9A	1.51	2.1
		11	4.31		^{11}A	0.19	2.88
Ni_4O	B	1	2.27	C_s	$^1A''$	2.17	1.69
		3	0.29	C_{2v}	$^3A_2$	1.40	2.45
		5	0.00	C_{2v}	$^5A_2$	0.84	2.05
		7	0.43	C_1	^7A	2.15	1.46
		9	1.48	C_s	$^9A''$	1.91	2.18
		11	3.06	C_{2v}	$^{11}A_1$	0.38	2.22
Ni_4O	C	1	2.87	C_2	^1A	3.66	1.57
		3	0.73	C_{3v}	$^3A_2$	3.45	2.72
		5	0.00	C_{2v}	$^5A_2$	4.71	2.44
		7	0.82	C_1	^7A	1.86	1.57
		9	1.78	C_{3v}	$^9A_1$	0.91	1.89
		11	4.16	C_s	$^{11}A''$	1.45	2.77
Ni_4O	D	1	1.98	C_{3v}	$^1A_1$	2.55	1.22
		3	1.2	C_1	^3A	3.98	2.01
		5	0.00	C_s	$^5A'$	2.18	2.2
		7	0.16	C_1	^7A	3.59	2.48
		9	1.16	C_s	$^9A'$	2.54	2.69
		11	3.57		$^{11}A''$	1.44	2.77
Ni_5O	A	1	3.05	C_s	$^1A'$	2.97	1.36
		3	0.15		$^3A''$	3.37	2.6
		5	0.16		$^5A'$	3.36	2.47
		7	0		$^7A''$	3.23	1.77
		9	0.68	C_1	^9A	2.30	1.91
		11	1.93		^{11}A	0.90	2.37
Ni_5O	B	1	2.99	C_1	^1A	2.28	1.49
		3	0	C_s	$^3A''$	3.37	2.45
		5	0.02		$^5A''$	3.30	2.49
		7	0.14		$^7A'$	3.12	1.74
		9	0.76	C_1	^9A	2.28	1.93
		11	3.00	C_s	$^{11}A'$	1.06	2.02
Ni_5O	C	1	3.38	C_s	$^1A''$	1.43	1.27
		3	0.03		$^3A'$	1.50	2.67
		5	0.00		$^5A'$	1.43	2.6
		7	0.22		$^7A''$	2.40	1.78
		9	0.63		$^9A''$	2.30	1.91
			2.23		$^{11}A'$	1.06	2.25
Ni_5O	D	1	2.99	C_s	$^1A''$	2.94	1.36
		3	0.32	C_4	^3B	1.77	2.47

Table 3 (continued)

Cluster	Isomer	Multiplicity	Relative energy (eV)	Symmetry	IR*	DM** (debye)	HOMO-LUMO gap (eV)
		5	0.26	C_{4v}	$^5A_2$	1.66	2.4
		7	0.00	C_s	$^7A''$	2.89	1.93
		9	1.27	C_4	^9B	0.91	1.83
		11	1.89	C_1	^{11}A	0.86	2.38
Ni_6O	A	1	4.19	C_s	$^1A'$	2.25	1.34
		3	0.44		$^3A''$	1.35	2.1
		5	0.63		$^5A''$	1.66	1.71
		7	0.00	C_{2v}	$^7B_2$	1.63	2.15
		9	0.61	C_s	$^9A''$	1.18	1.41
		11	1.41		$^{11}A'$	0.55	2.34
Ni_6O	B	1	0.33	C_1	^1A	2.97	2.35
		3	0.00		^3A	3.34	1.85
		5	0.26	C_s	$^5A''$	1.32	2.51
		7	0.44	C_1	^7A	3.14	1.51
		9	0.22		^9A	2.43	2.24
		11	1.14		^{11}A	1.23	1.94

* TR⇒Irreducible representation; ** DM⇒ Electrostatic dipole moment

initially and shows a relative maximum at n=4 in Ni_nO cluster. Beyond this point there is hardly any increase of BE per atom. However, the total optimized energy of the Ni_nO cluster can continue to gain energy during the growth processes. On the other hand, the binding energy in pure clusters tend to increase even if it is a cluster with n=7, which is basically indicating that clusters are gaining binding energy as well as the total optimized energy during their growth process. Comparing the binding energy curves it is clear that binding energy of the oxidized cluster is much higher than the nickel clusters of same size (n) within the range of the present calculation. This is an indication of improved stability due to the addition of oxygen in pure nickel structures.

The other parameters, which can explain the thermodynamic stability of the clusters, are embedding energy (EE), fragmentation energy (FE) and the second order difference in energy ($\Delta_2(n)$). The last parameter $\Delta_2(n)$ is also known as stability parameter. In the section calculated EE, FE and $\Delta_2(n)$ of Ni_nO and Ni_n clusters will be discussed. Variation of the embedding energy is shown in Fig. 4. Embedding energy is defined as,

$$EE = E(Ni_n) + E(O) - E(Ni_nO) \qquad (2)$$

It is actually the gain in energy to incorporate a foreign atom in a cluster during its growth process with an aim to produce a particular targeted product. In the present work since the growth process of Ni_nO studied by adding oxygen with Ni_n cluster, therefore, here EE is the gain in energy by the product cluster due to addition of oxygen with Ni_n.

Variation of embedding energy shown in Fig. 3 clearly shows that the cluster with n=4 in quintet state has an affinity to absorb oxygen atom to increase its stability compared to its surrounding clusters of same or different sizes. Or in other words, Ni_4 in quintet spin state is the most favorable size that can easily incorporate an oxygen atom in the cluster. A related but significantly related question is which is the most stable cluster as successive Ni atoms are added to increase the cluster size of Ni_nO starting from NiO. This is given by $\Delta_2(n)$ parameter or stability parameter of the clusters which is defined as,

$$\Delta_2(n) = \{E(Ni_{n+1}O) - E(Ni_nO)\} - \{E(Ni_nO) - E(Ni_{n-1}O)\}$$
$$= E(Ni_{n+1}O) + E(Ni_{n-1}O) - 2E(Ni_nO) \qquad (3)$$

With 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. In Fig. 4 calculated $\Delta_2(n)$ as a function of cluster size n is plotted. It is to be noted that $\Delta_2(n)$ exhibits a peak at n=4 which is the stability parameter of Ni_4O cluster in quintet state. Furthermore, the relative stability of Ni_4O (Fig. 4) cluster in terms of the calculated second order change in energy is the biggest among all the Ni_nO clusters in the present range of calculation showing that the cluster has stronger relative stability. Therefore, the structure with n=4 is the most stable geometry among all the members of Ni_nO (n=1-6) clusters.
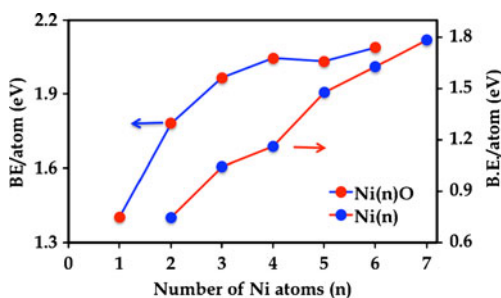
Fig. 2 Variation of binding energy per atom of Ni_n and Ni_nO (n=1–6) clusters with n

Fig. 4 Variation of stability of Ni_nO (n=1–6) clusters with n

Calculated fragmentation energies (Δ(n,n-1) or FE) of Ni_nO clusters can give the information of relative stabilities of the clusters. FE of the cluster is defined as,

$$\Delta(n, n-1) = E(Ni_nO) - \{E(Ni_{n-1}O) + E(Ni)\} \quad (4)$$

$$= E(Ni_nO) - E(Ni) - E(Ni_{n-1}O).$$

Variation of FE with the size of the cluster (Fig. 5) shows that there is a sharp increase of FE from n=3 to n=4 and then a drop from n=4 to 5 indicate that during the growth process of Ni_nO by adding Ni atom one by one the cluster Ni_4O is found most stable with quintet spin state.

While the above parameters indicate thermodynamic stability of a cluster, kinematic stability of the clusters in the chemical reaction is indicated by HOMO-LUMO gaps, ionization potential (IP), electron affinity (EA), chemical potential and hardness. Usually the larger the HOMO-LUMO gap, the less reactive the cluster is. HOMO-LUMO gaps of Ni_n and Ni_nO are plotted in Fig. 6. Overall, there is a decrease of gap with the increase of the cluster size in both Ni_n and Ni_nO clusters. However, there are some local oscillations over and above the decreasing trend. Although, there is no sharp global peak in HOMO-LUMO gap variation of Ni_nO cluster as in other quantities, but there is a visible local peak at n=4 in Ni_nO series. This again points out the enhanced stability at n=4 in Ni_nO series as like other quantities discussed before. This is precisely seen

in the variation of ionization potential and electron affinity with the cluster size shown in Figs. 7 and 8 respectively. Variation of the IP of Ni_nO first decreases from n=1 and the show a local maxima at n=4. Therefore, the cluster at n=4 has relatively higher chemical stability compare to the surrounding sizes. The sharp dip of electron affinity (EA) in Fig. 8 indicates the reaction by capturing electron from surrounding environment is relatively less in Ni_4O cluster compared to the other oxidized and pure nickel clusters. This also supports the enhanced stability of Ni_4O cluster. The common feature in all of these graphs at n=4, is indicating that with the size-evolution of Ni_nO clusters, the cluster Ni_4O in quintet spin state has an enhanced local stability.

Further, to verify the chemical stability of the 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+ΔN electrons where $\Delta N \ll N$ can be expressed as:

$$E(N + \Delta N) = E(N) + \left.\frac{dE}{dx}\right|_{x=N} \Delta N$$

$$+ \frac{1}{2}\left.\frac{d^2E}{dx^2}\right|_{x=N} (\Delta N)^2$$

$$+ \text{ neglected higher order terms.} \quad (5)$$

Fig. 3 Variation of embedding energy of Ni_nO (n=1–6) clusters with n
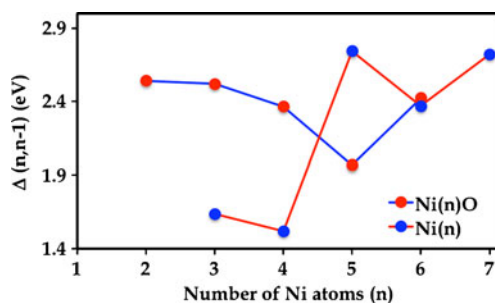
Fig. 5 Variation of fragmentation energy of Ni_nO (n=1–6) clusters with n
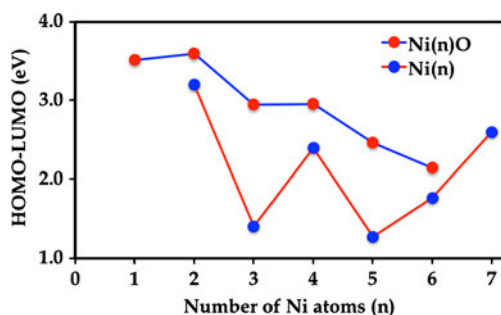
Fig. 6 Variation of HOMO-LUMO gap of Ni_n and Ni_nO (n=1–6) clusters with n

Fig. 8 Variation of electron affinity of Ni_nO (n=1–6) clusters with n

Then, μ and η can be defined as:

$$\mu = \left.\frac{dE}{dx}\right|_{x=N} \quad \text{and} \quad \eta = \left.\frac{1}{2}\frac{d^2E}{dx^2}\right|_{x=N} = \left.\frac{1}{2}\frac{d\mu}{dx}\right|_{x=N} \quad (6)$$

Since,

$$IP = E(N-1) - E(N) \text{ and } EA = E(N) - E(N+1). \quad (7)$$

By setting $\Delta N=1$, μ and η are related to IP and EA via the following relations:

$$\mu = -\frac{IP + EA}{2} \quad \text{and} \quad \eta = \frac{IP - EA}{2} \quad (8)$$

Consider two systems with μ_i and η_i (i=1,2) contracting each other, where some amount of electronic charge (ΔM) transfer from one to other. The quantity ΔM and the resultant energy change (ΔE) due to the charge transfer can be determined in the following way:

If E (N+ΔM) 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:

$$E_1(N_1 + \Delta M) = E_1(N_1) + \mu_1(\Delta M) + \eta_1(\Delta M)^2 \quad (9)$$

and

$$E_2(N_2 - \Delta M) = E_2(N_2) - \mu_2(\Delta M) + \eta_2(\Delta M)^2 \quad (10)$$

Corresponding chemical potential becomes,

$$\mu_1' = \left.\frac{dE_1(x + \Delta M)}{dx}\right|_{x=N_1} = \mu_1 + 2\eta_1\Delta M \text{ and} \quad (11)$$

$$\mu_2' = \left.\frac{dE_2(x - \Delta M)}{dx}\right|_{x=N_2} = \mu_2 - 2\eta_2\Delta M$$

to first order in ΔM after the charge transfer. In chemical equilibrium, $\mu_1' = \mu_2'$ which gives the following expressions:

$$\Delta M = \frac{\mu_2 - \mu_1}{2(\eta_1 + \eta_2)} \quad \text{and} \quad \Delta E = \frac{(\mu_2 - \mu_1)^2}{2(\eta_1 + \eta_2)}. \quad (12)$$

In the expression, ΔE is the energy gain by the total system (1 and 2) due solely to the alignment of chemical potential of the two systems at the same value. It is evident from the above expressions of ΔM and ΔE that a guiding principle to predict the occurrence of an easier charge transfer is a large difference in μ together with low η_1 and η_2.

Keeping these aspects in mind, chemical potential (μ) and chemical hardness (η) of the clusters are calculated. The size variations of chemical potential and chemical hardness are shown in Figs. 9 and 10 respectively. Variations of these parameters also support the higher stability of Ni_4O cluster in neutral state. Variation of chemical potential shows a relative minimum at n=4. It means that the reaction affinity of this cluster is relatively low compare to its surrounding sizes. Again the variation of chemical hardness, shows a stable minimum nature of all the sizes for n>2, which is another indication of non-reactive nature of the cluster. Therefore the nature of chemical potential and chemical hardness graphs support the highest stability of the neutral Ni_4O cluster.

To get an idea about the effect of the addition of oxygen in the pure nickel clusters HOMO-LUMO of the ground state structures at different sizes and geometries are shown by using contour mapping in Fig. 1b in addition to the resultant distribution of the total spin density. Comparing the maps of pure and doped clusters, it can be seen that due to addition of oxygen there is a drift of charge and spin densities toward the nickel atoms. Therefore, in the contour

Fig. 7 Variation of ionization potential of Ni_nO (n=1–6) clusters with n
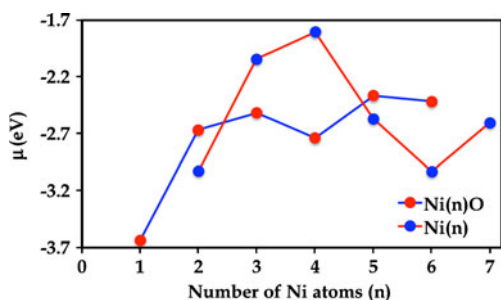
Fig. 9 Variation of chemical potential of Ni_nO ($n=1$–6) clusters with n

maps the wavefunction representing the HOMO orbital or in other words, electron density in the neighboring space of oxygen atoms are comparatively less. The change of electron density from HOMO-1 to HOMO and LUMO to LUMO+ 1 are also studied (not shown in figure) in ground state stable structures. It shows that the oxygen atom is able to attract only a weak orbital charge density in its surrounding space. Specially, in the most stable ground state structure, Ni_4O, the electron clouds mainly localizes around Ni atoms while the distribution of electron density around O atom is relatively low. This is indicating that the orbital hybridization between Ni and O atoms is relatively less than orbital hybridization between the Ni and Ni atoms in the cluster. Therefore, the shapes of the HOMO and bonding properties between bare Ni_n clusters and Ni_nO clusters are obviously different. Furthermore, the Ni-Ni bonding in the bare Ni_n clusters is the localized σ -type bonds or partly delocalized σ-type bonds while the bonding among Ni-Ni atoms in the Ni_nO clusters is completely delocalized σ-type bonds. Consequently, the hybridization between Ni and O atoms are responsible for the change of electronic properties of Ni_n clusters after adsorbing O atom and it helps in improving the stability of Ni_nO clusters because of high ionization potential, relatively high HOMO-LUMO gap, less electron affinity and chemical potential.

Variations of calculated magnetic moments of pure and doped clusters of the most stable structures are shown in Fig. 11. It is important to mention here that the average magnetic moment of the clusters changes discontinuously with the size of clusters. This behavior of the magnetic
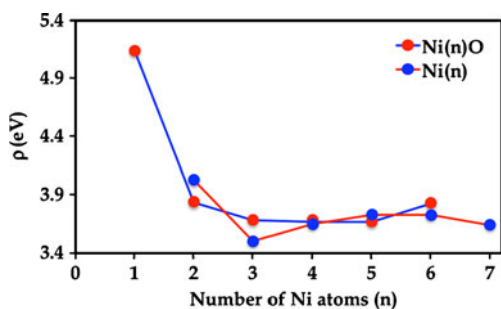
moments of the low energy small sized Ni_nO clusters is clearly indicating that the magnetic moments for the Ni_nO clusters are not stable in a particular size; it is very much geometry dependent and this finding is analogous to those of the reported Zr_n clusters [55] as well as in transition metal doped Na clusters [56]. Furthermore, the magnetic moment is stable or in other words it is independent of geometry of the stable isomers in Ni_4O. Therefore, the incoming oxygen does not have much effect on the magnetic moment of the ground state of Ni_4O clusters. But the question is how the ten (eight from four nickel and two from one oxygen atoms) electrons are accommodating in the Ni_4O ground state cluster so that the magnetic moment of Ni_4 cluster is not affected by addition of oxygen and at the same time increased the chemical stability? The answer can be given by Mulliken population charge and spin density analysis. Mulliken charge analysis clearly shows that out of ten electrons, six electrons take part in the bonding orbital's p_x, p_y and p_z. Out of four remaining electrons, two electrons go into the p_y^* orbital. One electron goes into the majority spin level p_z^* orbital, whereas the down spin level orbital p_z^* is empty. Since the nickel is strongly ferromagnetic element, so the remaining one electron goes into the lowest unoccupied spin minority level. Again from the spin density analysis it is found that the oxygen atom has positive magnetic moment in the Ni_4O cluster in quintet state. This is because of the splitting of sub-bands originating from oxygen $2p_y$ and $2p_z$ levels due to ferromagnetism of nickel clusters.

Conclusions

In the present study a systematic investigation of the Ni_n and Ni_nO ($n=1$-6) clusters in various spin states are performed at the unrestricted B3PW91/LanL2DZ level. Total optimized energies, equilibrium geometries, and stabilities of Ni_nO ($n=1$-6) clusters along with the binding energy, embedding energy, fragmentation energies, ionization potential, electron affinity etc. are presented and discussed. Conclusions drawn from the calculated results can be summarized as follows:

Fig. 10 Variation of chemical hardness of Ni_nO ($n=1$–6) clusters with n
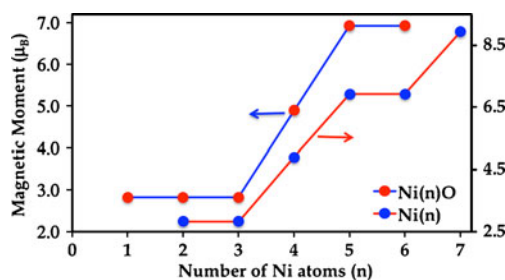
Fig. 11 Variation of magnetic moment of Ni_nO ($n=1$–6) clusters with n

(1) The ground state geometry in Ni_nO (n=1-6) clusters show that the O atom prefers the surface capped Ni_n pattern. The oxygen atom interacts with the three Ni atoms in the most stable Ni_4O cluster in the Ni_nO series in such a way that the stability of the cluster increases.

(2) The relative stabilities of Ni_nO (n=1-6) clusters in terms of different calculated parameters show that Ni_4O cluster is the most stable structure among all small size (n<7) nickel and nickel oxide clusters.

(3) Calculated magnetic moments of small-sized Ni_nO clusters show that the magnetic moment is not stable and it depends upon the geometrical structure of the cluster in a particular size. Whereas, in the most stable cluster Ni_4O in C_s symmetry structure has the magnetic moment of $5.0\,\mu B$, which is the same as the most stable pure Ni_4 cluster. At this size the magnetic moment is independent of the geometry of the cluster and also negligible effect of addition of oxygen on it.

(4) Calculated HOMO-LUMO gaps indicate that the most stable Ni_4O cluster in quintet spin state has stronger chemical stability as compared to the neighboring sizes. The distribution of electron density of the HOMO states for the most stable Ni_4O cluster mainly localizes around Ni atoms while the distribution around O atom is relatively low. The shapes of the HOMO and bonding properties between bare Ni_n clusters and Ni_nO clusters are obviously different. Our calculated results on electron affinity, chemical potential etc. indicate that the incoming oxygen does not have much effect on the magnetic property of Ni_n clusters, but at the same time it helps in improving the stability of the cluster after forming Ni_nO.

References

1. Veliah S, Xiang KH, Pandey R, Recio JM, Newsam JM (1998) J Phys Chem B 102:1126–1135
2. Cui XY, Morrison I, Han G (2002) J Chem Phys 117:1077–1084
3. Han JG, Ren ZY, Lu BZ (2004) J Phys Chem A 108:5100–5110
4. Molek KS, Jaeger TD, Duncan MA (2005) J Chem Phys 123:144313–144322
5. Andrews L, Chertihin GV, Ricca A, Bauschlicher CW Jr (1996) J Am Chem Soc 118:467–470
6. Wang LS, Wu H, Desai SR (1996) Phys Rev B 53:8028–8031
7. Tono K, Terasaki A, Ohta T (2006) J Chem Phys 124:184311–184316
8. Asmis KR, Sauer J (2007) Mass Spectrom Rev 26:542–562
9. Feyel S, Scharfenberg L, Daniel C, Hartl H, Schröder D, Schwarz H (2007) J Phys Chem A 111:3278–3286
10. Tsipis AC, Tsipis CA (2000) J Phys Chem A 104:859–865
11. Molek KS, Reed ZD, Ricks AM, Duncan MA (2007) J Phys Chem A 111:8080–8089
12. Li S, Dixon DA (2007) J Phys Chem A 111:11908–11921
13. Herman J, Foutch JD, Davico GE (2007) J Phys Chem A 111:2461–2468
14. Xu Y, Shelton WA, Schneider WF (2006) J Phys Chem A 110:5839–5846
15. Zhai HJ, Huang X, Waters T, Wang XB, O'Hair RAJ, Wedd AG, Wang JS (2005) J Phys Chem A 109(46):10512–10520
16. Liu L, Zhao RN, Han JG, Liu FY, Pan GQ, Sheng LS (2009) J Phys Chem A 113:360–366
17. Li S, Hennigan JM, Dixon DA, Peterson KA (2009) J Phys Chem A 113:7861–7877
18. Li S, Zhai HJ, Wang LS, Dixon DA (2009) J Phys Chem A 113:11273–11288
19. Rodríguez-Lopez JL, Aguilera-Granja F, Vega A, Alonso JA (1999) Eur Phys J D 6:235–241
20. Yao YH, Gu X, Ji M, Gong XG, Wang DS (2007) Phys Lett A 360:629–631
21. Grigoryan VG, Springborg M (2001) Phys Chem Chem Phys 3:5135–5139
22. Lathiotakis NN, Andriotis AN, Menon M, Connolly J (1996) J Chem Phys 104:992–1003
23. Reddy BV, Nayak SK, Khanna SN, Rao BK, Jena P (1998) J Phys Chem A 102:1748–1759
24. Czekaj I, Wambach J, Kröcher O (2009) Int J Mol Sci 10:4310–4329
25. Bahl CRH, Hansen MF, Pederse T, Saadi S, Nielsen KH, Lebech B, Mørup S (2006) J Phys Condens Matter 18:4161–4175
26. Khadar MA, Biju V, Inoue A (2003) Mater Res Bull 38:1341–1349
27. Ahmad T, Ramanujachary KV, Lofland SE, Ganguli AK (2006) Solid State Sci 8:425–430
28. Biesinger MC, Payne BP, Lau LWM, Gerson A, Smart RSC (2009) Surf Interface Anal 41:324–332
29. Ghosh M, Biswas K, Sundaresana A, Rao CNR (2006) J Mater Chem 16:106–111
30. Vann WD, Castleman AW Jr (1999) J Phys Chem A 103:847–857
31. Perdew JP, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Singh DJ, Fiolhais C (1992) Phys Rev B 46:6671–6678
32. Perdew JP, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Singh DJ, Fiolhais C (1993) Phys Rev B 48:4978
33. Perdew JP, Burke K, Wang Y (1996) Phys Rev B 54:16533–16539
34. Burke K, Perdew JP, Wang Y (1998) In: Dobson JF, Vignale G, Das MP (eds) Electronic density functional theory: recent progress and new directions. Plenum
35. Kohn W, Sham LJ (1965) Phys Rev 140A1:135–138
36. Langreth DC, Perdew JP (1980) Phys Rev B 21:5469–5493
37. Langreth DC, Mehl MJ (1983) Phys Rev B 28:1809–1834
38. Perdew JP, Wang Y (1986) Phys Rev B 33:8800–8802
39. Perdew JP (1986) Phys Rev B 33:8824
40. Bandyopadhyay D (2008) J Appl Phys 104:084308–084314
41. 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 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
42. Ho J, Polak ML, Ervin KM, Lineberger WC (1993) J Chem Phys 99:8542–8551

43. Wu H, Wang LS (1997) J Chem Phys 107:9606–9611
44. Sadanov VI, Harris DO (1988) J Chem Phys 89:2748–2753
45. Noell JO, Newton MD, Hay PJ, Martin RL, Bobrowicz FW (1980) J Chem Phys 73:2360–2371
46. Gutsev GL, Lutatskaya VD, Klyagina AP, Levin AA (1986) Chem Mater Sci 28:829–833
47. Doll K, Dolg M, Fulde P, Stoll H (1997) Phys Rev B 55:10282–10288
48. Bauschlicher CW, Maitre P (1995) Theor Chim Acta 90:189–203
49. Mitchell KAR (1985) Surf Secrets 149:93–104
50. Basch H, Newton MD, Moskowitz JW (1980) J Chem Phys 73:4492–4510
51. Nygren MA, Siegbahn PEM, Wahlgren U, Akeby H (1992) J Phys Chem 96:3633–3640
52. Tomonari M, Tatewaki H, Nakamura T (1986) J Chem Phys 85:2875–2884
53. Pastor GM, Dorantes-Davila J, Bennemann KH (1988) Chem Phys Lett 148:459–464
54. Blyholder G (1974) Surf Sci 42:249–260
55. Wang CC, Zhao RN, Han JG (2006) J Chem Phys 124:194301–194308
56. Pradhan K, Sen P, Reveles JU, Khanna SN (2008) Phys Rev B 77:045408–045414

ORIGINAL PAPER

Theoretical study of the hydroboration reaction of cyclopropane with borane

Satya Prakash Singh · Pompozhi Protasis Thankachan

Abstract The hydroboration of cyclopropane has been investigated using the B3LYP density functional method employing several split-valence basis sets. It is shown that the reaction proceeds via an intermediate weakly bound complex and a three-centered transition state. Calculations at higher levels of theory were also performed at the geometries optimized at the B3LYP level, but only slight changes in the barriers were observed. Structural parameters for the transition state are also reported.

Keywords Cyclopropane · Hydroboration · Intermediate complex · Propylborane · Transition state

Introduction

The parent member of the set of hydrides of boron having general formula BxHy, called boranes, is BH_3, which is itself called borane. It does not exist independently except transiently and dimerises to give diborane, B_2H_6 [1] and forms stable complexes with ether or amines. BH_3 adds to double bonds in steps; thus with ethylene $C_2H_5BH_2$, $(C_2H_5)_2BH$ and $(C_2H_5)_3B$ are formed. Treatment of $C_2H_5BH_2$ with hydrogen peroxide in alkaline medium yields ethanol. This sequence of hydroboration and oxidation converts alkenes to alcohols; it is extremely useful synthetically because of its selectivity. It results in *syn*-addition of water to an alkene (Scheme 1) and is an anti-Morkownikoff [2, 3] process where the hydroxy group attaches itself to the less substituted carbon. The synthetic utility of the reaction was originally discovered by Brown [4, 5].

Mechanism-wise, the addition of BH_3 to alkenes was generally understood to take place in a concerted manner, with simultaneous bond-making and bond-breaking events, but the calculations of Nagase et al. [6] and Wang et al. [7] showed that BH_3 forms a π-complex with ethylene, which then goes on to form the adduct; hence the reaction is really a two-step process. Further, they showed that the complex has a triangular geometry and that the transition state in the second step has a four-centered structure. On the other hand, Seyferth [8] and Streitwieser [9] on stereochemical grounds predicted a three-centered transition state. Later work by Hommes et al. [10] supports the three-centered transition state for alkene hydroboration.

Cyclopropane behaves like olefins in several instances; for example, it undergoes addition reactions [11]. Graham and stone [12] have shown that the gas-phase reaction of cyclopropane and diborane results in several products including tri-*n*-propyl borane. However, to our knowledge there have been no reports of theoretical studies of the hydroboration of cyclopropane. In the present paper, we have reported the results of our investigations of the reaction between cyclopropane and borane at DFT level of theory.

Computational methods

All calculations have been performed on a PC running WINDOWS using the Gaussian 98 [13] suite of programs. The B3LYP hybrid density functional was used for calculation at DFT [14] level using several split-valence basis sets. The geometries of the reactants (BH_3 and C_3H_6), intermediate complex (**LM1**), transition state (**TS**) and the

S. P. Singh · P. P. Thankachan (✉)
Department of Chemistry, Indian Institute of Technology Roorkee, Roorkee 247667, India
e-mail: thankachaniitr@gmail.com
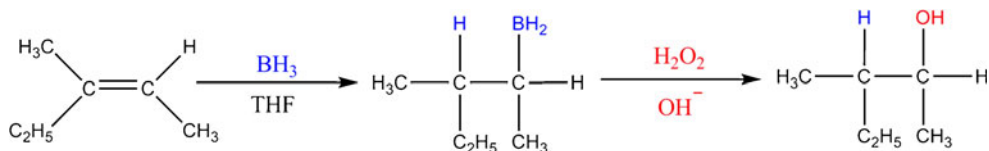
Scheme 1 Syn-addition of water to an alkene by hydroboration-oxidation

Fig. 1 B3LYP/6-31G** optimized geometries of the reactants (BH$_3$ and C$_3$H$_6$), complex (LM1), transition state (TS) and product (LM2). The energy values (in kcal mol^{-1}) relative to the reactants (BH$_3$ and C$_3$H$_6$) are shown in parenthesis
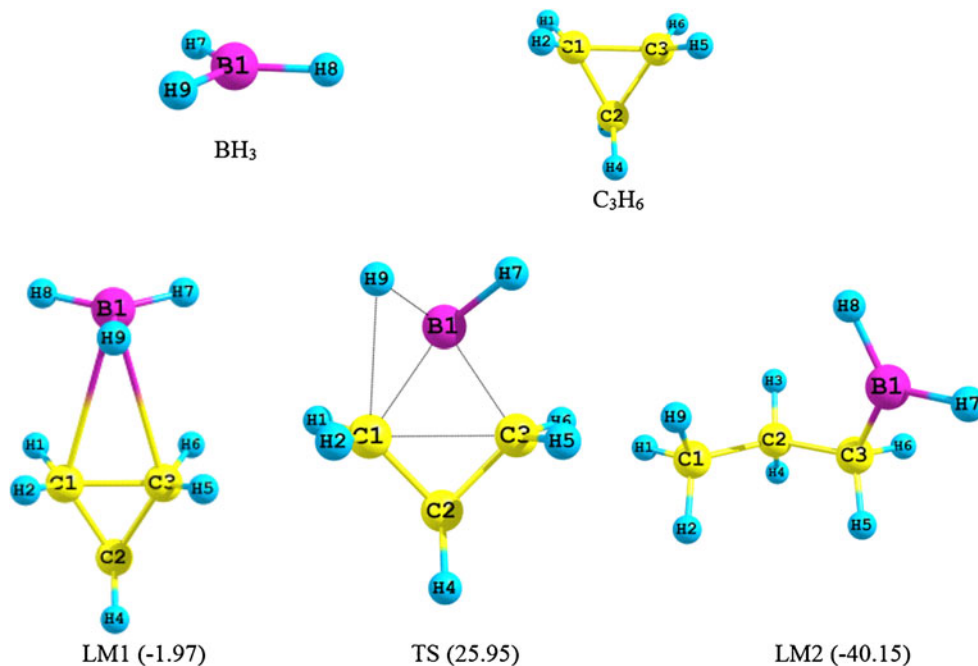
Table 1 Optimized structural parameters (bond lengths in Å and angles in degree) for the reactants, complex (LM1), transition state (TS) and product (LM2)

	BH$_3$	C$_3$H$_6$	LM1	TS	LM2
R(B1-H7)	1.19234	–	1.19194	1.21194	1.19723
R(B1-H8)	1.19233	–	1.19194	1.19976	1.19855
R(B1-H9)	1.19234	–	1.19271	1.21309	2.98101
R(C1-C3)	–	1.50829	1.52619	1.99438	2.55284
R(C2-C3)	–	1.50829	1.50946	1.50953	1.53748
R(C1-C2)	–	1.50829	1.50546	1.50457	1.53197
R(C1-H1)	–	1.08581	1.08468	1.08626	1.09484
R(C1-H2)	–	1.08581	1.08494	1.08674	1.09631
R(C2-H3)	–	1.08581	1.08562	1.50953	1.09745
R(C2-H4)	–	1.08581	1.08551	1.09208	1.09777
R(C3-H5)	–	1.08581	1.08494	1.08720	1.09999
R(C3-H6)	–	1.08581	1.08468	1.08722	1.11015
R(B1-C1)	–	–	2.90515	1.85288	3.28006
R(B1-C3)	–	–	2.90518	1.81343	1.55917
R(C1-H9)	–	–	3.14961	2.06138	1.09643
∠ C1C2C3	–	60.000	60.913	82.857	112.546
∠ C2C3C1	–	60.000	59.543	48.465	33.658
∠ C3C1C2	–	60.000	59.543	48.678	33.796
∠ H7B1H9	119.999	–	119.869	90.570	137.381
∠ H8B1H9	120.001	–	119.869	113.955	71.811
∠ H8B1H7	120.000	–	119.907	115.724	118.086
∠ C2C3B1	–	–	133.869	102.225	118.457
∠ C3B1H9	–	–	90.473	131.991	68.653

product (**LM2**) were optimized at this level of theory using 6-31G**, 6-31++G** [15–21], cc-pVDZ, cc-pVTZ and AUG-cc-pVTZ [22–26] basis sets. The nature of each stationary point was confirmed in each case by frequency calculations; all the minima were verified to have all positive frequencies and the transition state to have only one imaginary frequency. MP2 [27–31] calculations were also performed using the chosen basis sets for comparison purposes. Single point (SP) calculations were also performed at the geometries optimized at the B3LYP/6-31G**
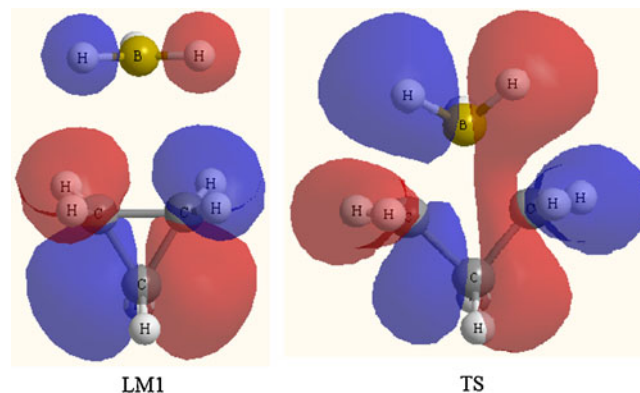
Fig. 2 HOMOs of complex (LM1) and transition state (TS) for the hydroboration of cyclopropane at B3LYP/6-31G** level of calculation
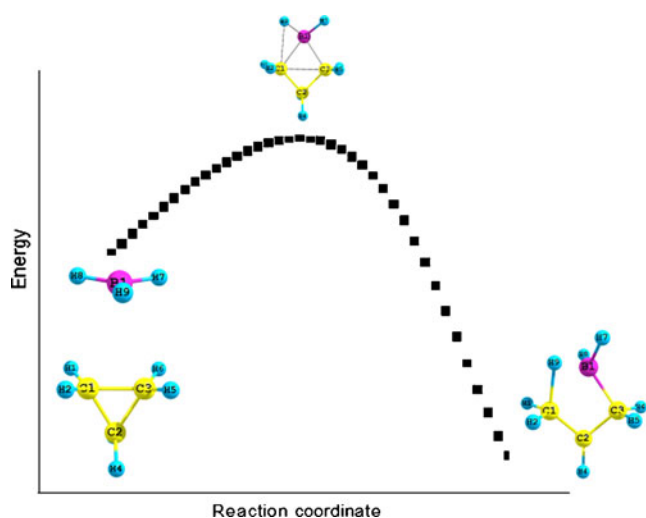
Fig. 3 IRC plot for the transition state at the B3LYP/6-31G** level

level at the CCSD, CCSD(T) [32–36], QCISD, QCISD(T) [36] and MP4D [22] levels to see if any significant change in the energetics is observed.

Results and discussion

The geometries of the reactants (BH₃ and C₃H₆), the intermediate complex (**LM1**), the transition state (**TS**) and product (**LM2**) optimized at the B3LYP/6-31G** level are shown in Fig. 1. The optimized geometrical parameters for these species are collected in Table 1. The predicted boron-carbon (B1-C1 and B1-C3) distances of 2.905 Å and the C1-C3 distance of 1.526 Å in the intermediate complex are

Table 2 Unscaled vibrational frequencies of reactants (BH₃ and C₃H₆), intermediate complex (LM1), transition structure (TS) and the addition product (LM2) at B3LYP/6-31G** level of calculation

Species	Normal modes of vibrations (cm⁻¹)
BH₃	1162, 1213, 1213, 2582, 2715, 2715
C₃H₆	739, 739, 862, 891, 892, 1065, 1065, 1088, 1157, 1217, 1218, 1223, 1486, 1486, 1535, 3140, 3140, 3147, 3217, 3217, 3237
LM1	47, 83, 106, 106, 363, 386, 756, 757, 858, 865, 893, 1053, 1075, 1093, 1140, 1167, 1208, 1209, 1215, 1216, 1223, 1485,1486, 1530, 2581, 2708, 2711, 3144, 3153, 3156, 3224, 3234, 3249
TS	390i, 84, 211, 410, 627, 662, 684, 752, 824, 962, 1002, 1029, 1052, 1083, 1142, 1154, 1176, 1202, 1245, 1269, 1295, 1461, 1477, 1519, 2453, 2463, 2593, 3083, 3123, 3132, 3140, 3210, 3222
LM2	125, 181, 231, 305, 392, 542, 776, 785, 905, 932, 981, 1056, 1099, 1126, 1223, 1250, 1289, 1365, 1377, 1417, 1429, 1505, 1514, 1521, 2588, 2658, 2935, 3028, 3031, 3034, 3063, 3101, 3001

Table 3 Total energies (in Hartree) at different levels and basis sets

Method	BH₃+C₃H₆	LM1	TS	LM2
B3LYP/6-31G**	−144.5194864	−144.5226268	−144.4781284	−144.5834724
B3LYP/cc-pVDZ	−144.5074419	−144.5097716	−144.4673293	−144.5722806
B3LYP/6-311++G**	−144.5520729	−144.5538850	−144.5088965	−144.6141527
B3LYP/AUG-cc-pVTZ	−144.5678575	−144.5695796	−144.5240755	−144.6284606
MP2/6-31G**	−143.9833023	−143.9884832	−143.9363884	−144.0464528
MP2/cc-pVDZ	−143.9742346	−143.9793004	−143.9321547	−144.0397117
MP2/6-311++G**	−144.0336348	−144.0387580	−143.9909528	−144.0968889
ᵃQCISD/6-31G**	−144.0404988	−144.0446920	−143.9863016	−144.1036544
ᵃQCISD(T)/6-31G**	−144.0559733	−144.0609200	−144.0078051	−144.1194505
ᵃMP4D/6-31G**	−144.0458756	−144.0503929	−143.9937177	−144.1090801
ᵃCCSD/6-31G**	−144.0399797	−144.0441433	−143.9852795	−144.1030577
ᵃCCSD(T)/6-31G**	−144.0557719	−144.0606962	−144.0074457	−144.1192105

ᵃ Single point calculations on the B3LYP/6-31G** optimized geometries

slightly higher than the values predicted for the corresponding distances (2.828, 2.835 and 1.322 Å respectively) in the case of the π-complex formed between BH₃ and ethylene during hydroboration of ethylene [6]. This suggests that the complex is weakly bound in comparison with the complex of borane with ethylene.

The transition state (**TS**) is a three-center one. The distances corresponding to forming bonds (B1-C3 and C1-H9, Table 1) are smaller than in the intermediate complex (**LM1**). The B1C3C1H9 core in the transition state (**TS**) is nearly planar in contrast to the nearly perpendicular arrangement of the substrate relative to the B1-H9 axis found in intermediate complex (**LM1**). The angle between the line joining the boron (B1) to the midpoint of the C1-C3 bond and the breaking B1-H9 bond in the transition

Table 4 Relative energies (in kcal mol⁻¹) at different levels and basis sets

Method	Reactants	ΔE_LM 1	ΔE_TS	ΔE_LM 2
B3LYP/6-31G**	0.00	−1.97	25.95	−40.15
B3LYP/cc-pVDZ	0.00	−1.46	25.17	−40.68
B3LYP/6-311++G**	0.00	−1.14	27.09	−38.95
B3LYP/AUG-cc-pVTZ	0.00	−1.08	27.47	−31.30
MP2/6-31G**	0.00	−3.25	29.43	−39.62
MP2/cc-pVDZ	0.00	−3.18	26.40	−41.08
MP2/6-311++G**	0.00	−3.21	26.78	−39.69
ᵃQCISD/6-31G**	0.00	−2.63	34.00	−39.63
ᵃQCISD(T)/6-31G**	0.00	−3.10	30.22	−39.83
ᵃMP4D/6-31G**	0.00	−2.83	32.72	−39.66
ᵃCCSD/6-31G**	0.00	−2.61	34.32	−39.58
ᵃCCSD(T)/6-31G**	0.00	−3.09	30.32	−39.80

ᵃ Single point calculations on the B3LYP/6-31G** structures

structure (**TS**) is nearly 88.5 degree, very close to the expected value of 90 degree for a three-centered structure [37]. The formation of the product propylborane (**LM2**) from the transition state (**TS**) involves H9, originally attached to the boron (B1) shifting closer to C1 and the boron (B1) moving toward C3, the bond between C1 and C3 lengthening and ultimately breaking. The HOMOs of the transition structure and the intermediate complex are shown in Fig. 2 for visualization of electron density distribution. It is seen that the electron distribution of cyclopropane is largely intact in the complex, whereas there is a visible degradation in the transition structure. Similar changes have been reported for the π-bound ethylene-alane system [38].

IRC calculations have been performed and confirm that the transition structure (**TS**) does fall on the path between the complex (**LM1**) and the product (**LM2**). The IRC plot shown in Fig. 3 shows the transition structure (**TS**) moving downhill toward the complex (**LM1**) on one side and the product (**LM2**) the other. Frequency calculations have been performed on all stationary structures, and the results are shown in Table 2. All positive frequencies are found for the stable species (BH_3, C_3H_6, LM1 and LM2) and one imaginary frequency (390 cm^{-1}) is observed in the case of TS.

The calculations have been performed also using cc-pVDZ, 6-311++G** and AUG-cc-pVTZ basis sets at B3LYP level, and at MP2 level using 6-31G**, cc-pVDZ and 6-311++G** basis sets. Single point calculations at QCISD, QCISD(T), MP4D, CCSD and CCSD(T) levels have also been performed on the structures optimized at the B3LYP level using the 6-31G** basis set. The total energies (in Hartree) and the relative energies (in kcal mol^{-1}) obtained are summarized in Table 3 and Table 4 respectively.

The qualitative trends found strictly parallel those found in the B3LYP/6-31G** calculations, thus confirming the suitability of this DFT scheme for similar studies. We also note that the MP2 energies of the complex (**LM1**) are somewhat lower than in the other cases and the values show less variability with the basis set chosen. These results suggest an energy barrier less than 30.00 kcal mol^{-1}, making the reaction viable; however, no experimental results are available for comparison.

Concluding remarks

In summary, we have studied the stationary structures involved in the hydroboration of cyclopropane with borane. Our study posits a three-centered transition state for this reaction in contrast to most of the earlier studies on the hydroboration of alkenes, which reported four-centered transition state. It is also hoped that studies on reactions involving cyclopropane and its derivatives with other hydroboration reagents will clarify the situation.

References

1. Schlesinger HI, Burg AB (1942) Chem Rev 31:1–41
2. Brown HC, Zweifel G (1960) J Am Chem Soc 82:4708–4712
3. Olah GA, Hockswender TR Jr (1974) J Org Chem 39:3478–3481
4. Brown HC, Benjamin WA (1962) Hydroboration. New York
5. Brown HC (1972) Boranes in organic chemistry. Cornell University Press, Ithaca
6. Nagase S, Ray NK, Morokuma K (1980) J Am Chem Soc 102:4536–4537
7. Wang X, Li Y, Wu YD, Paddon-Row MN, Rondan NG, Houk KN (1990) J Org Chem 55:2601–2609
8. Seyferth D (1962) In: Cotton FA (ed) Progress in Inorganic Chemistry, vol 3. Wiley, New York, p 210
9. Streitwieser A, Verbit L, Bittman R (1967) J Org Chem 32:1530–1532
10. Hommes VE, Nicolaas JR, Schleyer PvR (1991) J Org Chem 56:4074–4076
11. Cromwell NH, Graff MA (1952) J Org Chem 17:414–425
12. Graham WAG, Stone FGA (1957) Chem Ind (London) 1096–1097
13. Frisch MJ et al. (1998) Gaussian98, Revision A.7. Gaussian Inc, Pittsburgh PA
14. Becke AD (1993) J Chem Phys 98:5648–5652
15. Ditchfield R, Hehre WJ, Pople JA (1971) J Chem Phys 54:724–728
16. Hehre WJ, Ditchfield R, Pople JA (1972) J Chem Phys 56:2257–2261
17. Hariharan PC, Pople JA (1974) Mol Phys 27:209–213
18. Gordon MS (1980) Chem Phys Lett 76:163–168
19. Hariharan PC, Pople JA (1972) Theor Chim Acta 28:213–222
20. Clark T, Chandrasekhar J, Spitznagel GW, Schleyer PvR (1983) J Comput Chem 4:294–301
21. Frisch MJ, Pople JA, Binkley JS (1984) J Chem Phys 80:3265–3269
22. Woon DE, Dunning TH (1993) J Chem Phys 98:1358–1371
23. Kendall RA, Dunning TH, Harrison RJ (1992) J Chem Phys 96:6796–6806
24. Dunning TH Jr (1989) J Chem Phys 90:1007–1023
25. Peterson KA, Woon DE, Dunning TH Jr (1994) J Chem Phys 100:7410–7415
26. Wilson A, Mourik Tv, Dunning TH Jr (1996) J Mol Struct THEOCHEM 388:339–349
27. Head-Gordon M, Pople JA, Frisch MJ (1988) Chem Phys Lett 153:503–506
28. Frisch MJ, Head-Gordon M, Pople JA (1990) Chem Phys Lett 166:275–280
29. Frisch MJ, Head-Gordon M, Pople JA (1990) Chem Phys Lett 166:281–289
30. Head-Gordon M, Head-Gordon T (1994) Chem Phys Lett 220:122–128
31. Saebo S, Almlof J (1989) Chem Phys Lett 154:83–89
32. Cizek J (1969) Adv Chem Phys 14:35–89
33. Purvis GD, Bartlett RJ (1982) J Chem Phys 76:1910–1918
34. Scuseria GE, Janssen CL, Schaefer HF (1988) J Chem Phys 89:7382–7387
35. Scuseria GE, Schaefer HF (1989) J Chem Phys 90:3700–3703
36. Pople JA, Head-Gordon M, Raghavachari K (1987) J Chem Phys 87:5968–5975
37. This angle in a perfect four-center structure with all bonds the same length would be 63 degree
38. Bundens JW, Yudenfreund J, Francl MM (1999) Organomet 18:3913–3920

ORIGINAL PAPER

Identification of new potential *Mycobacterium tuberculosis* shikimate kinase inhibitors through molecular docking simulations

Carolina Pasa Vianna · Walter F. de Azevedo Jr.

Abstract Tuberculosis (TB) is the major cause of human mortality from a curable infectious disease, attacking mainly in developing countries. Among targets identified in *Mycobacterium tuberculosis* genome, enzymes of the shikimate pathway deserve special attention, since they are essential to the survival of the microorganism and absent in mammals. The object of our study is shikimate kinase (SK), the fifth enzyme of this pathway. We applied virtual screening methods in order to identify new potential inhibitors for this enzyme. In this work we employed MOLDOCK program in all molecular docking simulations. Accuracy of enzyme-ligand docking was validated on a set of 12 SK-ligand complexes for which crystallographic structures were available, generating root-mean square deviations below 2.0 Å. Application of this protocol against a commercially available database allowed identification of new molecules with potential to become drugs against TB. Besides, we have identified the binding cavity residues that are essential to intermolecular interactions of this enzyme.

Keywords Molecular docking · Shikimate kinase · Shikimate pathway · Tuberculosis · Virtual screening

C. P. Vianna · W. F. de Azevedo Jr. (✉)
Faculdade de Biociências, Instituto Nacional de Ciência
e Tecnologia em Tuberculose (INCT-TB), Laboratório de
Bioquímica Estrutural (LaBioQuest), Pontifícia Universidade
Católica do Rio Grande do Sul (PUCRS),
Av. Ipiranga 6681,
Porto Alegre, RS 90619-900, Brazil
e-mail: walter@azevedolab.net

C. P. Vianna · W. F. de Azevedo Jr.
Programa de Pós Graduação em Biologia Celular e Molecular,
Pontifícia Universidade Católica do Rio Grande do Sul,
Porto Alegre, RS, Brazil

Introduction

Tuberculosis (TB) is the most important cause of human death from a curable infectious disease. It is estimated that, worldwide, one hundred million people are infected annually and about ten million develop the disease, with five million of those progressing to an infectious stage, culminating with approximately three million deaths. According to the World Health Organization [1], the overall incidence of TB increases approximately 0.3% per year. The resurgence of this health problem occurred mainly due to the proliferation of multi (MDR-TB), extensively (XDR-TB), and recently, totally-drug (TDR-TB) resistant Mt strains. Besides, the high susceptibility of HIV/AIDS infected patients to TB is also a health problem. Therefore, there is an urgent need for the discovery and development of new and better drugs for the TB treatment [2].

Enzymes of the shikimate pathway (SP) are promising targets for the development of antimicrobial agents [3] and herbicides [4], because they are essential to the survival of algae, higher plants, bacteria, fungi, apicomplexan parasites and absent in mammals [5]. It is a seven-step biosynthetic route that converts erythrose 4-phosphate to chorismate, a precursor of aromatic amino acids and many other essential compounds [6].

The object of our study is the fifth enzyme of the SP, shikimate kinase (SK) (EC 2.7.1.71), which catalyzes the specific phosphorylation of the 3-hydroxy group of shikimate using ATP as a co-substrate resulting in shikimate-3-phosphate and ADP [7, 8]. This enzyme is an established target against Mt, since Parish and Stoeker demonstrated that the SP is essential for the viability of Mt due to the disruption of the aroK gene, which codes for the SK enzyme [9].

SK is a member of the nucleoside monophosphate kinases (NMP kinases) family, which suffer large conformational changes during catalysis (Fig. 1) [10]. The enzymes of this family are composed of three domains: the CORE, which contains a highly conserved phosphate-binding loop (P-loop), the LID domain, which undergoes substantial structural changes upon substrate binding, and the NMP-binding domain which is responsible for the recognition and binding of a specific substrate [11].

Drugs are usually discovered by trial and error by means of high-throughput screening approaches that use in vitro experiments to evaluate the activity of a large number of compounds against a known target. This procedure is very costly and time-consuming. If crystallographic information is available for the protein target, then molecular docking simulations can be a helpful computational approach in the drug-discovery process [12]. Molecular docking is a simulation method that predicts the conformation of a receptor-ligand complex, in which the receptor can be either a protein or a nucleic acid, and the ligand is a small molecule. This computer simulation can generate many possible positions for the ligand in the receptor-binding pocket. Therefore, a criterion is necessary that will allow comparisons of all possible positions of the ligand, and then a selection can be made for the best position.

Our goal here is to find potential inhibitors against shikimate kinase from *Mycobacterium tuberculosis* MtSK using virtual screening (VS). VS can decrease costs and improve hits rates for lead discovery. For this, we used the MtSK structure [13] as a target for the molecular docking simulations with MOLDOCK [14]. Our docking protocol was validated against an ensemble of 12 crystallographic structures available for complexes of MtSK. The VS was validated by inclusion of a known SK inhibitor in the small-molecule database with over 4500 structures. We describe the results obtained in terms of the MOLDOCK scores, modes of interaction and discuss the importance of the active site residues in the ligand binding process.

Materials and methods

Molecular docking simulations

One of the fundamental questions in structural biology is the study of protein-ligand interactions, particularly considering the pharmacological applications of such study in the design of drugs based on structure [15]. To simulate the interaction of MtSK with a library of ligands, we used the MOLDOCK program [14], an implementation of a variant of the evolutionary algorithm (EA). Recent evaluation of MOLDOCK strongly indicates that it is capable of finding the right position of a ligand. Furthermore, MOLDOCK exhibits better overall performance compared with SURFLEX, FLEXX, and GOLD [14]. In the present work, all simulations were performed in an iMac (Intel Processor Core 2 Duo, 2.66 GHz, 2 GB SDRAM DDR3 1066 MHz).

Re-docking and cross-docking

In molecular docking simulations, the best binary complex (protein-ligand) is the one closer to the crystallographic structure. For that reason we must establish a methodology that assesses the distance from the computer-generated solution (pose) to the crystallographic structure. This distance can be calculated using the root-mean-square deviation (RMSD), which is a measure of the differences between values predicted by a model and the values actually observed from the object being modeled or estimated (protein-ligand complex). The RMSD is calculated between two sets of atomic coordinates, in this case, one for the crystallographic structure (x_{ctal}, y_{ctal}, z_{ctal}; the object being modeled) and another for the atomic coordinates obtained from the docking simulations (x_{pose}, y_{pose}, z_{pose}; predicted model). A summation is then taken over all N atoms being compared, using the following equation:

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(x_{ctal,i} - x_{pose,i}\right)^2 + \left(y_{ctal,i} - y_{pose,i}\right)^2 + \left(z_{ctal,i} - z_{pose,i}\right)^2}. \tag{1}$$

In docking simulations, it is expected that the best results generate RMSD values less than 2.0 Å compared with crystallographic structures [16]. This procedure of obtaining the crystallographic position of the ligand is often called "re-docking," which is fundamentally a validation method that determines whether the molecular docking algorithm is able to recover the crystallographic position using computer

simulation. In this work, all RMSD calculations were calculated for non-hydrogen atoms.

In order to validate our docking protocol, we used the SK crystallographic coordinates available at the protein data bank (PDB), under the access code 2DFN [13]. We performed the docking simulation against the active site of MtSK and compared the docked poses with the crystallographic
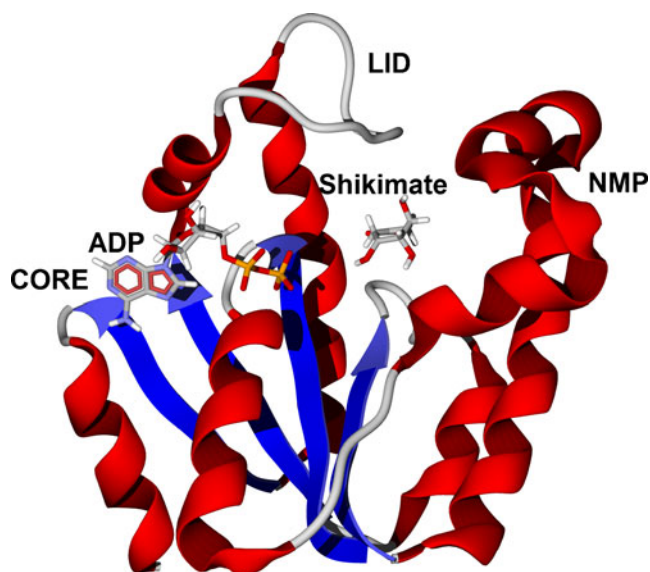
Fig. 1 Structure of shikimate kinase in complex with ADP and shikimate (PDB access code: 2DFN)

structure. We used the MOLDOCK default protocol with center at coordinates x=(−15.23), y=(−14.38), and z= (14.88) Å, and a docking sphere radius of 9 Å. Figure 2 shows the docking sphere used in the simulations.

In the implementation of EA in MOLDOCK, computational approximations of an evolution course, called genetic operators, are applied to simulate the permanence of the most positive features. In a sample space, where there is a problem or a search routine and many different possible solutions (candidates), each option is ranked based on a set of parameters (scoring function or fitness function), and only the best ranked solutions are kept for the next iteration. This cycle is repeated until an optimal solution can be found. In the molecular docking simulations, the optimal solution is the one with the best scoring function, which should be the closest to the crystallographic structure. MOLDOCK presents two biological inspired algorithms to perform positional searches in docking simulations. One is
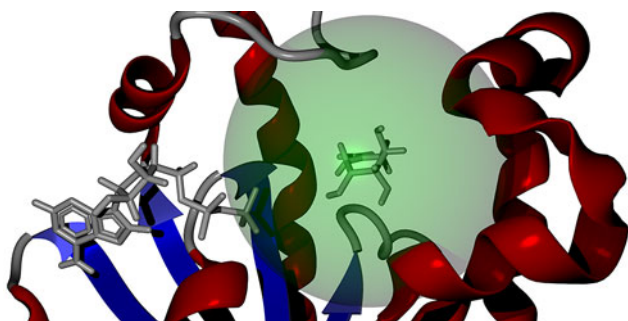
Fig. 2 Search space sphere (green) defined for molecular docking simulations

called the optimizer search algorithm (MOLDOCK Optimizer), which is based on an GDEA [14]. The second is a simplex evolution algorithm (SE) called MOLDOCK SE. GDEA is based on an EA adjustment called differential evolution (DE), which provides a distinct method for selecting and modifying candidate solutions (individuals). We used MOLDOCK Optimizer as search algorithm.

In addition to re-docking, a procedure called "cross-docking" can also be used to further validate a docking protocol. Considering that several crystallographic structures are available for the same protein, cross-docking can be applied. This procedure involves docking a number of ligands found in a variety of crystal structures of a protein identical to a single rigid protein crystallographic conformation [17]. When a protein target presents major conformational changes upon ligand binding, a significant difference is expected between the crystallographic and docked structures. We identified 12 MtSK structures in PDB with ligands in the shikimate-binding site (PDB access codes: 2DFN, 1U8A, 1WE2, 1ZYU, 2G1K, 2IYQ, 2IYR, 2IYS, 2IYX, 2IYY, 2IYZ, and 3BAF). This search was performed on February, 24th 2011. This validation procedures, re-docking and cross-docking, is the initial stage of a virtual screening protocol (**phase 1**) described in the next sections.

Virtual screening

Our virtual screening (VS) protocol is divided in four phases as shown in Fig. 3. **Phase 1** is focused on selection and validation of a docking protocol, as described earlier in the section re-docking and cross-docking. **Phase 1** ends when an adequate protocol is found (selection criterion RMSD<2.0 Å). It should be pointed out that the RMSD criterion is dependent on the number of torsion angles, and a less demanding criterion may be adopted for re-docking of a ligand with a number of torsion angles higher than 10 [14]. Once a docking protocol is chosen we select a small-molecule database to be used in the screening (**phase 2**). Here we used a ligand library commercially obtainable at Acros Organics. The ligands (mol2 format) were downloaded from http://zinc.dock.org [18], with a total of 4579 small molecules. In addition to commercially oriented databases the ZINC database also provides an interface to build small-molecule databases based on molecular similarity, such as Tanimoto coefficient [19, 20].

In **phase 3**, we start docking simulations for each ligand present in the selected database. MOLDOCK program is the workhorse of the present protocol. It was used in all docking simulations described here. During a typical docking simulation several orientations can be obtained for each ligand. Here we selected the one with the lowest

Fig. 3 Flowchart of virtual screening process

Phase 1

Phase 2

Phase 3

Phase 4
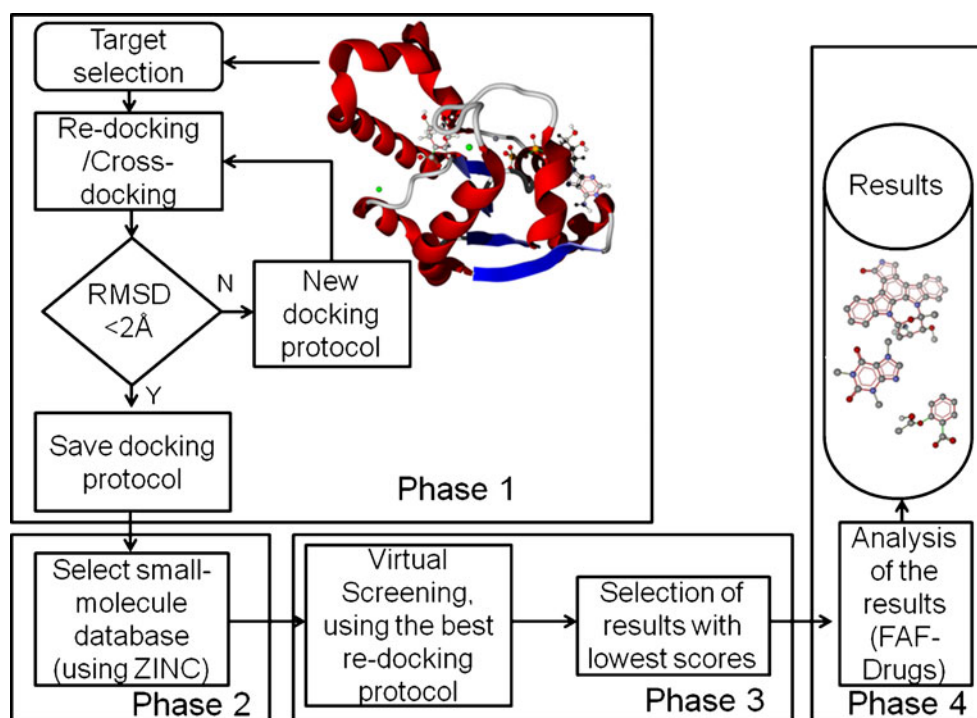
scoring function. The scoring function used by MOLDOCK improves accuracy of scoring functions with a new hydrogen bonding term and new charge schemes. Four scoring functions are implemented in the MOLDOCK, including MOLDOCK score and PLANTS score [14, 21]. These two functions offer grid-based versions, in which hydrogen bond directionality is not considered. In the present protocol we employed grid-based MOLDOCK score since it offers approximately four-fold greater speed by performing a precalculation of potential-energy values on an equally spaced cubic grid.

The MOLDOCK score is based on the piecewise linear potential (PLP) scoring functions developed by Yang et al. [22, 23]. The docking scoring function $E_{MOLDOCK\ SCORE}$ is defined as the following:

$$E_{MOLDOCK\ SCORE} = E_{Intramol} + E_{intermol}, \qquad (2)$$

where $E_{intermol}$ is the intermolecular interaction energy:

$$E_{intermol} = \sum_{i \in ligand} \sum_{j \in protein} \left[332 \frac{q_i q_j}{Dr_{ij}} + E_{PLP}(r_{ij}) \right]. \qquad (3)$$

All non-hydrogen atoms in the ligand and protein are taken in the summation. The first term accounts for electrostatic interactions, in which the factor 332 is used to obtain energy in kJ mol^{-1}. D represents the dielectric constant, which is the following: $D=4r_{ij}$. The second term (E_{PLP}) is a PLP, described elsewhere [22, 23]. To ensure that no energy term can be superior to the clash penalty, the electrostatic term is cut off at a level equivalent to the distance of 2.0 Å for distances less than 2.0 Å.

Intramolecular energy is given by the following equation:

$$E_{intramol} = E_{penalty} + \sum_{i \in ligand} \sum_{j \in ligand} E_{PLP}(r_{ij})$$
$$+ \sum_{singlebonds} A[1 - \cos(n\varphi - \varphi_0)]. \qquad (4)$$

The term $E_{penalty}$ is a penalty energy to be added to $E_{intramol}$ when two non-bonded atoms are closer than 2 Å (for non-hydrogen atoms). This term avoids unrealistic molecular topologies for the ligands. The second term is a PLP, already mentioned [22, 23]. The last term accounts for torsion energy, which is expressed as a periodic function. In this term, A, n, and ϕ_o are empirically determined [22, 23]. MOLDOCK defines a limiting sphere where the search is focused. If a ligand non-hydrogen atom is positioned outside this limiting sphere (the search space sphere), then a constant penalty of 10000 is added to the total energy (implemented for the grid-based version of the MOLDOCK score).

After identification of potential inhibitors by molecular docking simulations, the best scored ligands were submitted to the web server FAF-Drugs [24], in order to assess physical-chemical properties (**phase 4**). These are key properties that need to be considered in early stages of the drug discovery process, and FAF-Drugs allows users to filter molecules via simple rules such as molecular weight, polar surface area, logP and number of rotatable bonds. The ligands were filtered following the Lipinski's rule of five (RO5). RO5 advocates that drugs which present oral

bioavailability, in general, follow: molecular weight less or equal to 500, Log$P$ less or equal to 5, number of hydrogen bond donor groups less or equal to 5 and number of hydrogen bond acceptor groups less or equal to 10 [25].

In addition to the 4579 small molecules present in the Acros database we added staurosporine (PubChem Compound Identification: CID 44259) to the database to be used in the VS. This molecule has been already tested directly on MtSK [26]. This addition allows testing whether this VS protocol is able to identify SK inhibitors present in a database with over 4,500 ligands.

Enrichment factor

In order to further validate the present VS protocol we calculated the enrichment factor (EF), which takes into account the improvement of the hit rate by a VS protocol compared to a random selection. EF is defined by the following equation,

$$EF = \frac{H_a/H_t}{A/N} \qquad (5)$$

where H_a is the number of active compounds in the H_t top-ranked compounds of a total database of N compounds of which A are active [27, 28]. Successfully VS implies EF $\gg 1$. For this validation simulation, a sub-set of a previously described kinase decoy database (containing 627 molecules) was spiked with the four known MtSK inhibitors (A=4) [26]. This database for MtSK decoys and active MtSK inhibitors is available for downloading at: http://azevedolab.dominiotemporario.com/doc/mtsk_decoys_set.zip.

Results and discussion

Docking and cross-docking

Re-docking simulations (phase 1 of the VS protocol) using the structure 2DFN generated an RMSD of 1.6 Å. In addition, cross-docking simulations generated RMSD ranging from 1 to 2 Å further validating the present docking protocol. These two tests indicated that the docking simulation was successful, and that the protocol is good enough to be used for the virtual screening process.

Virtual screening

VS uses computational methodologies to identify biologically active molecules against a specific protein target. Two main methodologies are used in VS. Methods that search for similarity to validated ligands and molecular docking methods that require the use of crystallographic information of the target. Here we made use of the second approach. VS studies performed by other research groups have been previously published on the identification of MtSK inhibitors as antitubercular drugs by similar molecular docking procedures [29, 30]. Nevertheless, very limited information (experimental data) on the direct effect of these compounds on MtSK is yet available. The novelty of the present work relays on the method used in the VS and the selection of compounds according to their pharmacological properties. The VS simulations were carried out using the MOLDOCK program, having as target the MtSK (PDB access code 2DFN). The ligand library comprises 4580 molecules (Across database plus staurosporine). Addition of a known SK inhibitor allows testing the accuracy of the present protocol.

After docking simulations, we selected 20 top-scoring compounds from the initial set of 4580 compounds (selection based on MOLDOCK score). Staurosporine was present in the 20 top-scoring compounds obtained in the VS, with MOLDOCK score of −144.168. Identification of a known SK inhibitor among the best VS results gives further validation for this VS protocol. These 20 potential inhibitors were submitted to filter tests, available at the web server FAF-Drugs [24], to exclude those compounds that have known undesirable physical-chemical features to oral bioavailability. We could have applied this filter analysis previous to docking simulations, since it would reduce simulation time. Nevertheless, we kept filtering analysis after docking simulation, since the MOLDOCK protocol was fast enough to be run in less than a week of CPU time of an iMac (Intel Processor Core 2 Duo, 2.66 GHz, 2 GB SDRAM DDR3 1066 MHz). In addition, application of filtering analysis previous to docking simulations could eliminate candidates that fail to filtering analysis but present promising MOLDOCK score, which could have toxicity reduced by small modification in the structure.

Especially interesting is the fact that staurosporine is a well-known cyclin-dependent kinase (CDK) inhibitor that has a plethora of structural and functional studies [31–35]. Staurosporine is non-selective and too toxic for use in therapy, but UCN-01, a hydroxylated form of staurosporine (7-hydroxystaurosporine), shows greater selectivity for CDK and is currently undergoing clinical trials in the United States and Japan [33]. This opens new possibilities to test new molecular moieties as potential SK inhibitors, the CDK inhibitors that have already shown low toxic effects make a promising dataset to be explored as potential SK inhibitors.

The FAF-Drugs parameters used were those of the Lipinski's role of five [25]. From the set with 20 selected molecules, nine fit the Lipinski's role of five, which includes stauorosporine. Figures 4a–i show the molecular

Fig. 4 Molecular structures of the top-scoring compounds identified in the VS protocol. **a)** Staurosporine. **b)** ZINC15707201. **c)** ZINC20462780. **d)** ZINC15707234. **e)** ZINC15675581. **f)** ZINC15707188. **g)** ZINC22936889. **h)** ZINC20464408. **i)** ZINC22936937
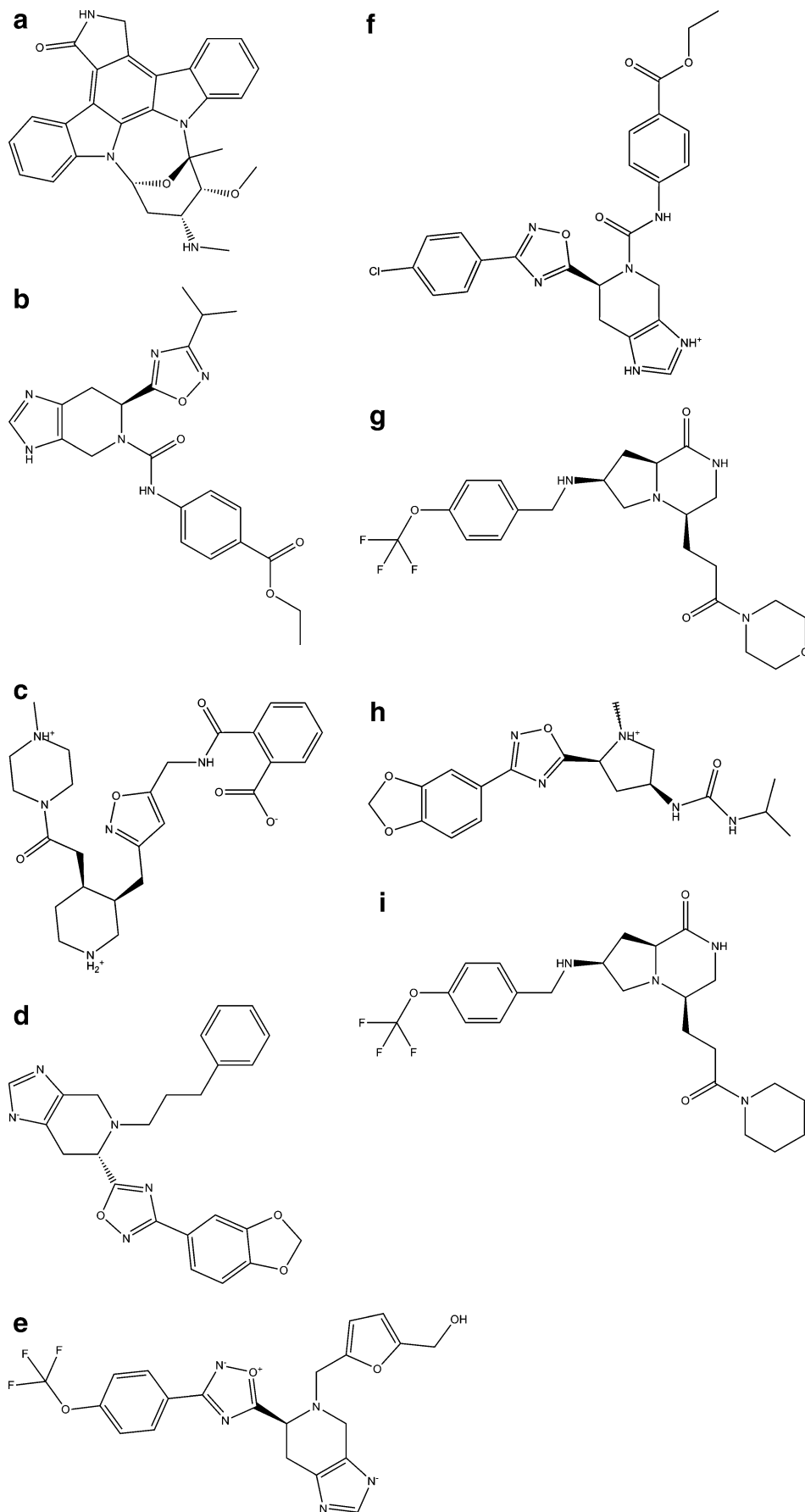
Table 1 Physical-chemical properties of ligands that fitted the Lipinki's role of five after analysis by FAF-Drugs

Ligand	ZINC code	Molecular weight (Da)	Number of HB acceptors	Number of HB donnors	LogP
1	15707201	423.3	10	1	0.82
2	20462780	485.3	10	4	-0.3
3	15707234	428.3	8	0	2.62
4	15675581	460.3	9	2	1.57
5	15707188	490.8	10	1	2.61
6	22936889	471.3	8	3	1.27
7	20464408	373.2	9	2	1.5
8	22936937	469.3	7	3	2.54

structures for all nine ligands. Since staurosporine is already a known SK inhibitor we excluded it from the rest of the analysis. Staurosporine was included only to test the VS protocol. The selected ligands are shown in Table 1. The MOLDOCK scores for these eight molecules ranging from −144.208 to −151.943. All eight ligands show MOLDOCK scores better than staurosporine.

Enrichment factor

The complete database used to calculate the EF contains 627 decoys and four active molecules (N=631). After application of the present VS protocol against this database, six molecules (H_t=6) were retrieved as hits (0.95% of the database). Among these hits, two molecules were known MtSK inhibitors (H_a=2) [26]. Thus, the enrichment factor was found to be 52.6, indicating that it is 52.6 times more likely to pick an active compound from the database than an inactive one. Previously published benchmarking sets for molecular docking of kinases exhibited EF ranging from 1.2 to 54, indicating that the present VS protocol is adequate for our purposes.

Intermolecular interactions

In order to better understand the interactions of these eight molecules with MtSK, we used the program LIGPLOT [36] to access the atoms of both, the small molecules and the protein ones that are responsible to make hydrogen bonds and van der Waals contacts. A comparison among the MOLDOCK score values obtained for these ligands, is not enough yet to predict activity, since in vitro assays are necessary to conclude this. Therefore it is not possible to say that the selected compounds, the ones with the best MOLDOCK scores, would be the most potent ones. We could observe, only, that among the selected compounds the best scores mean a greater potential to interact with the shikimate-binding cavity.

The docking simulation results corroborate the importance of some shikimate-active site residues as responsible

to establish intermolecular interactions with the substrate as well as with the tested ligands. The binding of shikimate to its cavity, presents pivotal residues that make protein-ligand interactions possible, as shown in Fig. 5. These residues are essential to the ligand binding and, finally, to the reaction catalyzed by the enzyme. The SK residues that perform intermolecular hydrogen bonds (HB) with the shikimate are: Gly80, Arg136 and Arg58. SK makes van der Waal contacts with residues: Ile45, Asp34, Pro11, Pro118, Gly79, Phe57, Leu119 and Gly81.

Information about intermolecular interactions for all eight top-scoring compounds is summarized in Table 2. Analysis of shikimate-binding site indicated that all top-scoring compounds present interaction with residues Lys15, Ser16 and Arg117. Figure 6 shows the intermolecular interactions for the top-scoring compound (ligand 1, ZINC15707201). This figure is representative of the positioning of all top-scoring compounds in the shikimate-binding pocket. Ligands 1, 2, 4, 5, 6 and 8 highlight the presence of residue Val116, suggesting it is also relevant to
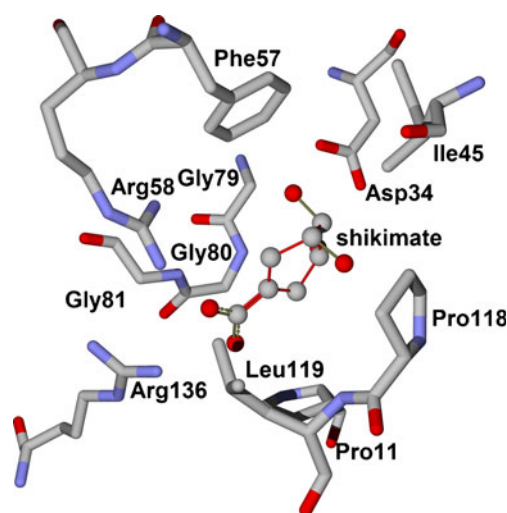
Fig. 5 Shikimate-binding pocket with main residues found in intermolecular interactions with shikimate

Table 2 Intermolecular interactions for the top-scoring ligands selected in the VS procedure. The presence of an X indicates that the interaction occurs. HB means hydrogen bonds and VDW means van der Waals contacts

Resídues	Ligands							
HB	Zinc15675581	Zinc15707188	Zinc15707201	Zinc15707234	Zinc20462780	Zinc20464408	Zinc22936889	Zinc22936937
Gly12						X		
Ser13						X		
Gly14						X		
Lys15	X	X	X	X	X	X		
Ser16		X	X					X
Thr17	X					X	X	X
Asp32		X	X					
Asp34	X						X	X
Leu78		X						
Arg117	X	X		X	X	X		
Leu119		X	X	X		X	X	
VDW								
Leu10					X			
Gly12	X	X	X	X		X		
Gly14	X	X	X	X		X	X	X
Lys15							X	X
Ser16	X			X	X	X	X	
Thr17		X	X	X	X			
Val35	X							
Gly38	X			X				
Phe49	X	X	X	X		X		X
Arg58			X					X
Gly80					X			
Val116	X	X	X		X		X	X
Arg117			X				X	X

intermolecular interactions. Two previously published molecular docking studies focused on MtSK were able to identify intermolecular molecular interactions with the same residues [29, 30], further corroborating the pivotal importance of these residues for ligand-binding affinity. Especially interesting is the fact that these previous molecular docking studies analyzed completely different molecular moieties, such as dipeptides (arginine-aspartate/

Fig. 6 Shikimate-binding pocket with main residues found in intermolecular interactions with the top-scoring compound (ZINC15707201)
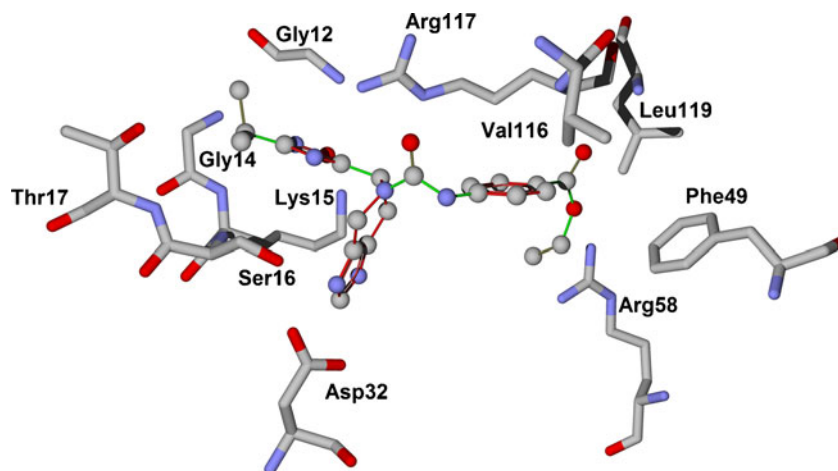
lysine-aspartate) [30] and triazole/tetrazole heteroaromatic systems [29]. These molecular structures were not present in the database used in the present study.

Conclusions

Advanced molecular docking algorithms available nowadays make it possible to undertake larger virtual screening studies focused on small-molecules libraries up to millions of compounds. Here we described an efficient molecular docking protocol, which was able to recover crystallographic position of a ligand present in the active site of the SK. Re-docking and cross-docking simulations generated RMSD results below 2 Å. The virtual screening protocol was able to confirm a known SK inhibitor, staurosporine, as a top-scoring compound and presents an enrichment factor of 52.6. Furthermore, the present work indicates new molecules with the potential to become drugs against TB. Besides, we identified the MtSK binding-cavity residues that are essential to make possible the interactions of this enzyme with a variety of molecules. Analysis of the top-scoring compounds also indicates that MtSK has the ability to bind a variety of molecular moieties not previously identified.

References

1. World Health Organization (2009) WHO Report. Geneva. Switzerland. WHO/CDS/TB/2009. http://www.who.int/topics/tuberculosis/en/. Accessed in October 2010
2. Basso LA, Pereira Da Silva LH, Fett-Neto AG, De Azevedo Jr WF, Moreira IS, Palma MS, Calixto JB, Astolfi Filho S, dos Santos RR, Soares MBP, Santos DS (2005) Mem Inst Oswaldo Cruz 100:475–506
3. Davies GM, Barrett-Bee KJ, Jude DA, Lehan M, Nichols WW, Pinder PE, Thain JL, Watkins WJ, Wilson RG (1994) Antimicrob Agents Chemother 38:403–406
4. Coggins JR (1989) The shikimate pathway as a target for herbicides. In: Dodge AD (ed) Herbicides and Plant metabolism. Cambridge University Press, Cambridge, UK, pp 97–112
5. Bentley R (1990) The shikimate pathway-a metabolic tree with many branches. Crit Rev Biochem Mol Biol 25:307–384
6. Ratledge C (1982) Nutrition, growth and metabolism. In: Ratledge C, Stanford JL (eds) The Biology of the Mycobacteria, vol 1. Academic, London, pp 185–271
7. Krell T, Coggins JR, Lapthorn AJ (1998) The three-dimensional structure of shikimate kinase. J Mol Biol 278:983–997
8. Pereira JH, Oliveira JS, Canduri F, Dias MVB, Palma MS, Basso LA, Santos DS, De Azevedo Jr WF (2004) Acta Crystallogr D Biol Crystallogr 60:2310–2319
9. Parish T, Stoker NG (2002) The common aromatic amino acid biosynthesis pathway is essential in Mycobacterium tuberculosis. Microbiology 148:3069–3077
10. Vonrhein C, Schlauderer GJ, Schulz GE (1995) Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. Structure 3:483–490
11. Gu Y, Reshetnikova L, Li Y, Wu Y, Yan H, Singh S, Ji X (2002) Crystal structure of shikimate kinase from mycobacterium tuberculosis reveals the dynamic role of the LID domain in catalysis. J Mol Biol 319:779–789
12. Peitsch MC (2004) Manuel Peitsch discusses knowledge management and informatics in drug discovery. Drug Discov Today Biocilico 02:94–96
13. Dias MV, Vasconcelos IB, Prado AM, Fadel V, Basso LA, De Azevedo Jr WF, Santos DS (2007) Acta Crystallogr F Struct Biol Cryst Commun 63:1–6
14. Thomsen R, Christensen MH (2006) MolDock: a new technique for high-accuracy molecular docking. J Med Chem 11:3315–3321
15. De Azevedo Jr WF, Canduri F, Da Silveira NJF (2002) Biochem Biophys Res Commun 293:566–571
16. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shaw DE, Shelley M, Perry JK, 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
17. Thilagavathi R, Mancera RL (2010) Ligand-protein cross-docking with water molecules. J Chem Inf Model 50:415–421
18. Irwin JJ, Shoichet BK (2005) ZINC - a free database of commercially available compounds for virtual screening. J Chem Inf Model 45:177–182
19. Timmers LFS, Pauli I, Caceres RA, De Azevedo Jr WF (2008) Drug-binding databases. Curr Drug Targets 9:1092–1099
20. De Azevedo Jr WF (2010) Structure-based virtual screening. Curr Drug Targets 11:261–263
21. De Azevedo Jr WF (2010) MolDock applied to structure-based virtual screening. Curr Drug Targets 11:327–334
22. Yang JM (2004) Development and evaluation of a generic evolutionary method for protein-ligand docking. J Comput Chem 25:843–857
23. Yang JM, Chen CC (2004) GEMDOCK: a generic evolutionary method for molecular docking. Proteins 55:288–304
24. Miteva MA, Violas S, Montes M, Gomez D, Tuffery P, Villoutreix BO (2006) FAF-drugs: free adme/tox filtering of compound collections. Nucleic Acids Res 34:W738–W744
25. 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
26. Mulabagal V, Calderón AI (2010) Development of an ultrafiltration-liquid chromatography/mass spectrometry (UF-LC/MS) based ligand-binding assay and an LC/MS based functional assay for Mycobacterium tuberculosis shikimate kinase. Anal Chem 82:3616–3621
27. Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. J Med Chem 49:6789–801
28. Brooijmans N, Sharp KA, Kuntz ID (2003) Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct 32:335–373
29. Segura-Cabrera A, Rodriguez-Perez MA (2008) Structure-based prediction of Mycobacterium tuberculosis shikimate kinase inhibitors by high-throughput virtual screening. Bioorg Med Chem Lett 18:3152–3157

30. Kumar M, Verma S, Sharma S, Srinivasan A, Singh TP, Kaur P (2010) Structure-based in silico design of a high-affinity dipeptide inhibitor for novel protein drug target Shikimate kinase of Mycobacterium tuberculosis. Chem Biol Drug Des 76:277–284

31. Noble ME, Endicott JA, Johnson LN (2004) Protein kinase inhibitors: insights into drug design from structure. Science 303:1800–1805

32. Canduri F, De Azevedo Jr WF (2005) Structural basis for interaction of inhibitors with cyclin-dependent kinase 2. Curr Comput-Aid Drug 1:53–64

33. Lawrie AM, Noble ME, Tunnah P, Brown NR, Johnson LN, Endicott JA (1997) Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with CDK2. Nat Struct Biol 4:796–801

34. De Azevedo Jr WF, Mueller-Dieckmann HJ, Schulze-Gahmen U, Worland PJ, Sausville E, Kim SH (1996) Structural basis for specificity and potency of a flavonoid inhibitor of human CDK2, a cell cycle kinase. Proc Natl Acad Sci USA 93:2735–2740

35. De Azevedo Jr WF, Leclerc S, Meijer L, Havlicek L, Strnad M, Kim SH (1997) Inhibition of cyclin-dependent kinases by purine analogues: crystal structure of human cdk2 complexed with roscovitine. Eur J Biochem 243:518–526

36. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Prot Eng 8:127–134

ORIGINAL PAPER

A MP2 and DFT study of the influence of complexation on the aromatic character of phosphole

Angeles Peña-Gallego · Jesús Rodríguez-Otero ·
Enrique M. Cabaleiro-Lago

Abstract This work is focused in three topical subjects: intermolecular interactions, metal ions, and aromaticity. A comprehensive MP2/6-31+G* and B3LYP/6-31+G* study of the influence of cation-π interactions on the aromatic character of phosphole was conducted. For this purpose, the structures of complexes were optimized at both theoretical level and different magnetic properties were evaluated. The main conclusion is the increase of the aromatic character of the phosphole when complexes with Li^+, Be^{2+}, and Al^{3+} are formed.

Keywords Aromaticity · Cation-π interactions · DFT calculations · Magnetic properties · MP2 calculations · Phospholes

Introduction

Intermolecular interactions with aromatic rings are crucial in a variety of chemical and biological processes. The study of fundamental intermolecular interactions and new types of interaction is essential for the design of drugs and other functional materials as well as for understanding cluster formation. The interaction involving aromatic systems is usually one of the following three types: cation-π, π-π or X-H-π.

The importance of metal ions is now an irrefutable reality. So their role has been clearly established in the regulation of enzymatic activity, and in function, folding, and stability of biological systems [1–4]. The interest of cation-π interactions was recognized in biological systems, and in the design of organic nanotubes or biological receptor models [5, 6]. Recently, different experimental and theoretical studies have tried to understand the cation-π interactions. The main conclusion of these studies was that the cation-π interactions are the strongest noncovalent interactions with high complexation energies [7–10]. Many of these studies were carried out on aromatic systems as benzene, pyrrole and aromatic amino acid residues [11–14].

This work is a computational study of the interaction between Li^+, Be^{2+}, and Al^{3+} with phosphole. While about a dozen crystal structures with Li^+ and Na^+ bounded to phosphorus systems are known [15–17], the metal ion binding to phospholes has not been widely studied. In this sense, only a computational study has been found. In this recent study, the authors have analyzed the σ and π complexation energies of various heteroaromatic systems with Li^+, Na^+, K^+, Mg^{2+}, and Ca^{2+} [18]. Their main objectives were the relative strength of σ and π binding modes and the regioselectivity of metal ion binding. The authors found that the strongest π complex among the five-membered heteroatomic cycles was that of pyrrole with all the metals except with Mg^{2+}.

The backbone of classical heterocyclic chemistry is fundamentally formed by five-membered heterocycles such as pyrroles, furans, and thiophenes together with six-membered pyridines. Their chemistry has been continuously developed reaching huge proportions. These molecules have an essential role in biological chemistry and a recent interest in applied chemistry (for example, doped films of polypyrrole have electroconducting properties). However, the chemistry

A. Peña-Gallego (✉) · J. Rodríguez-Otero
Departamento de Química Física, Facultade de Química,
Universidade de Santiago de Compostela,
Avda. das Ciencias s/n,
15782 Santiago de Compostela, Spain
e-mail: angeles.pena@usc.es

E. M. Cabaleiro-Lago
Departamento de Química Física, Facultade de Ciencias,
Universidade de Santiago de Compostela, Campus de Lugo,
Avda. Alfonso X El Sabio s/n,
27002 Lugo, Spain

of phospholes has been underdeveloped when compared to its nitrogen, oxygen, and sulfur counterparts. In fact, the first phosphole was discovered as late as 1959 [19, 20].

The aromaticity of the phosphole and its derivatives has been one of their most analysed properties [21]. The potential aromaticity of phosphole was discussed after its first practically applicable synthesis [22, 23] was reported in the 1960s. In different reviews the nonaromatic behavior of phospole was indicated [24–26]. This fact was explained in basis to the pyramidal preference of tricoordinate phosphorous in its compounds [27].

The planarity of the tricoordinate phosphorous can be influenced by substituents. In this sense, several cases have been studied: π-acceptor groups either at phosphorus or at the neighboring carbon [28, 29], bulky substituents at phosphorus [30–32], etc.

So, the effect of the complexation with metal cations on the aromaticity of phosphole seems a very interesting aim. As the aromaticity of phospholes has been interpreted in function of the pyramidality of the corresponding compound [21], the effect of complexation in this pyramidality have also been analyzed. These two important facts have not been studied in previous works.

Several studies have analyzed the possibility of increasing the aromaticity of phosphole with the inclusion of substituents that planarizes its structure. So, the effect of π-acceptor groups either at phosphorus or at the neighboring carbon have been studied [29, 33]. In this sense, the 1,2-BH_2-5-NH_2-phosphole has been employed in this work in order to compare the effect of this type of substituents and the complexation with cations.

Originally, the concept of aromaticity [34] was introduced to rationalize the structure, stability, and reactivity of benzene and related organic compounds. However, in the last decades, the boundaries of aromaticity have been extended to new areas of chemistry and to a large list of highly diverse species, which were unimaginable some decades ago. Although very useful, "delocalization" and "aromaticity" are virtual concepts, which cannot be measured directly experimentally. Consequently, many evaluation and visualization methods have been devised. In this paper, we have employed the magnetic properties in order to evaluate the aromatic character.

It is well known that aromatization affects magnetic properties such as magnetic susceptibility and its anisotropy, leading to especially negative values for such properties [35–37]. These magnitudes are global properties, which can be affected by parts of the molecule not directly implicated in the aromaticity. To avoid this problem, the nucleus independent chemical shift (NICS) proposed by Schleyer et al. is very useful. The NICS is defined as the negative of the magnetic shielding [38] and it can be evaluated at any point of the molecule, exhibiting very negatives values in the center of aromatic rings.

The anisotropy of the current-induced density (ACID) is a new method based on magnetic properties and developed by Herges and Geuenich [39]. This method allows the visualization of the ring current formed when a magnetic field is applied and allows us to study the electronic delocalization in molecules [40, 41].

Computational details

Geometries were optimized at the density functional theory (specifically, the Becke3LYP functional [42, 43]) level or Möller-Plesset perturbation level with the inclusion of energy corrections through second-order (MP2)

Fig. 1 The studied structures obtained at B3LYP level
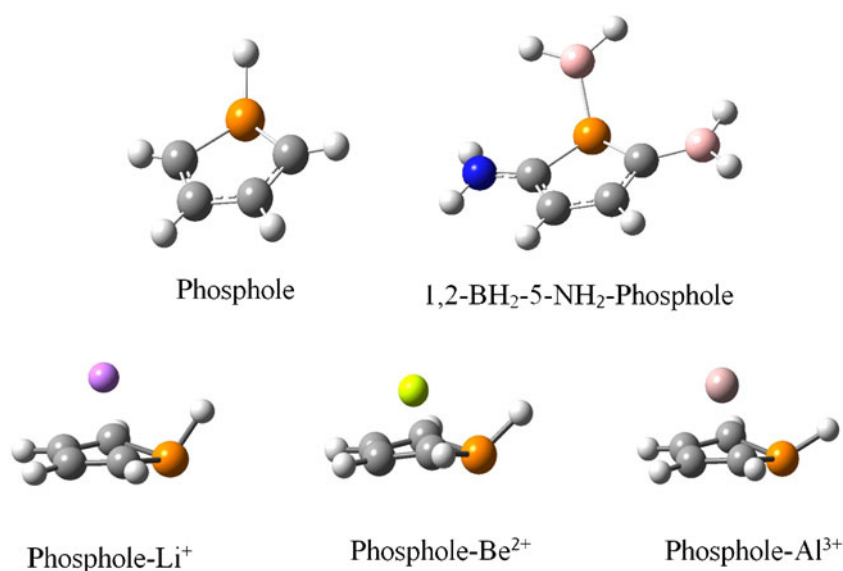
Table 1 Pyramidality in degrees, inversion barriers in kcal mol^{-1} (the values in parenthesis are corrected with ZPE) and distance cation-center in Å for the studied structures

		Phosphole	Li$^+$-phosphole		Be^{2+}-phosphole		Al^{3+}-phosphole		1,2-BH$_2$-5-NH$_2$-phosphole
ΣCPX	B3LYP	292.9	301.8		320.1		332.8		332.5
	MP2		307.7		327.9		337.7		330.9
Inversion barrier	B3LYP	18.65 (18.16)	11.09 (10.65)		7.57 (6.94)		3.28 (2.80)		0.90 (0.95)
	MP2	16.47 (15.76)	8.63 (8.15)		7.16 (6.43)		4.20 (3.65)		1.86 (1.48)
Distance cation-center			pyram.	planar	pyram.	planar	pyram.	planar	
	B3LYP		1.966	1.975	1.398	1.445	1.757	1.780	
	MP2		1.998	2.014	1.392	1.434	1.750	1.769	

with the 6-31+G* basis set. All points were confirmed as minima or transition states by calculating the harmonic vibrational frequencies at B3LYP/6-31+G* and MP2/6-31+G* levels, using analytical second derivatives. The pyramidal structures turned out to be minima and the planar structures are transition states (except for the phosphole-Al^{3+} planar structure which is a minimum at MP2 theory level).

The different possibilities for the complexes are considered (metal cation joined to phosphorous atom and cation-π cloud interaction in the two faces of the molecule).

The magnetic properties, nucleus-independent chemical shift (NICS), magnetic susceptibility (χ) and magnetic susceptibility anisotropy (χ$_{anis}$) were calculated. In the magnetic susceptibility calculations, the NMR shielding tensors have been computed with a larger basis set (6-311+G(2d,p)). In order to obtain the NICS at B3LYP/6-31+G* level, we have employed the gauge-independent atomic orbital (GIAO) method [44] but this method does not provide information about magnetic susceptibility, so χ, and χ$_{anis}$ were calculated using the IGAIM (Individual Gauges for Atoms in Molecules) method [45, 46], which is a slight variation of the continuous set of gauge transformations (CSCT) method [45–47]. Finally, CSGT method at B3LYP/6-31+G* level of theory was employed in ACID calculations, carried out with the program supplied by Herges and Geuenich [39].

All calculations were performed with the Gaussian98 [48] and Gassian03 [49] software packages.

Results and discussion

Structures

The different possibilities for the interaction between phosphole and the three metal cations have been analyzed. In all cases the more stable complexes were those obtained by the interaction between the π cloud of the phosphole and the cations in agreement with previous work [15–17]. For this interaction the two faces of the phosphole are accessible but the most favorable interaction is with the face where the hydrogen atom is placed (Fig. 1).

When the aromatic character of phospholes is analyzed, one of the most important points is the inversion barrier. This is one of the reasons why the planar structures for these complexes were studied. In all cases these structures turned out to be transition states as in the phosphole case, except for the phosphole-Al^{3+} planar structure which is a minimum at MP2 theory level.

The other reason to obtain the planar structures is based on an intrinsic problem when aromaticity is discussed. The magnetic properties (mean susceptibility, anisotropy of the susceptibility and nucleus independent shift, NICS) are used in order to indicate the aromatic character. High negatives values for these properties means aromatic character, but a model system is necessary in order to decide how high (aromatic) the value is. So, we think that the comparison of the values of these properties in the complexes and in their planar structures may be a good idea.

Table 2 Magnetic properties for the DFT-optimized structures

	Phosphole		Li$^+$-phosphole		Be^{2+}-phosphole		Al^{3+}-phosphole		1,2-BH$_2$-5-NH$_2$-phosphole	
	pyram.	planar	pyram.	planar	pyram.	planar	pyram.	planar	pyram.	planar
χ	-51.23	-59.42	-54.48	-60.60	-51.85	-53.35	-31.27	-24.73	-66.12	-68.32
χ$_{anis}$	-42.04	-52.84	-44.23	-49.20	-53.90	-51.84	-88.04	-96.93	-51.56	-54.60
NICS (0)	-5.44	-16.66	-7.52	-15.24	-13.35	-15.42	4.80	10.41	-6.04	-6.24
NICS (1)	-5.62	-10.60	-7.44	-10.42	-9.88	-10.07	-4.94	-3.21	-5.49	-5.29

Table 3 Magnetic properties for the MP2-optimized structures

	Phosphole		Li$^+$-phosphole		Be^{2+}-phosphole		Al^{3+}-phosphole	
	pyram.	planar	pyram.	planar	pyram.	planar	pyram.	planar
χ	-51.69	-59.47	-55.03	-60.46	-52.30	-53.35	-30.97	-25.54
χ_{anis}	-43.76	-53.22	-46.03	-49.47	-54.19	-51.92	-89.92	-96.31
NICS (0)	-6.29	-16.84	-9.62	-16.80	-14.82	-19.37	3.64	8.75
NICS (1)	-6.57	-11.15	-8.83	-10.79	-10.51	-12.06	-3.97	-2.50

Table 1 shows the pyramidality at the tricoordinate phosphorous atom (measured by the sum of the bond angles around the P-H group ΣCPX), the inversion barrier and distance of the metal cation to center of phosphole molecule in the analyzed structures at B3LYP/6-31+G* and MP2/6-31+G* levels of theory. The results at B3LYP and MP2 level are quite similar: the distances metal cation-center of phosphole are practically the same; the MP2 structures are, in general, more planar than the B3LYP structures but this fact is general when structures at DFT and MP2 level are compared; the inversion barriers follow the same pattern.

Two points are necessary to stand out: a) the pyramidality of phosphole is reduced with complexation and b) the inversion barrier is also reduced. In both cases the effect increases from Li$^+$ to Al^{3+}. In this table, the results for the 1,2-BH$_2$-5-NH$_2$-phosphole are also presented as comparison. In this case, with a planarity very similar to Al^{3+}-phosphole, the inversion barrier is rather low. This point indicates the importance of taking into account that not only pyramidality has an influence on the study of the aromaticity of phosphole and derivatives.

With regard to distances between center of phosphole moiety and cations, the longer values are for Li$^+$ cation and the shorter distances are for Be^{2+}. In all cases, the distances in the pyramidal and planar structures are very similar.

Magnetic properties and aromaticity

Susceptibility, anisotropy of susceptibility and NICS are calculated for phosphole and complex with cations. It is well know that aromatization affects magnetic properties such as magnetic susceptibility and its anisotropy, leading to specially

negative values for such properties [35–37]. The obtained values for these properties are collected in Table 2 (values for the DFT-optimized structures) and Table 3 (values for the MP2-optimized structures). The NICS values are obtained in the geometric center of the phosphole moiety. NICS values at points in the ring plane (NICS(0)) contain important spurious contributions from the in-plane tensor components that are not related to aromaticity as Schleyer has indicated [50]. For this reason, NICS(1) and NICS(−1) values (1 Å above/below the plane of the ring) are shown in Table 1. These values reflect π effects and they are a better indicator of the ring current than values in ring plane, because at 1 Å

Fig. 2 Figure indicating the position where NICS(1) is evaluated
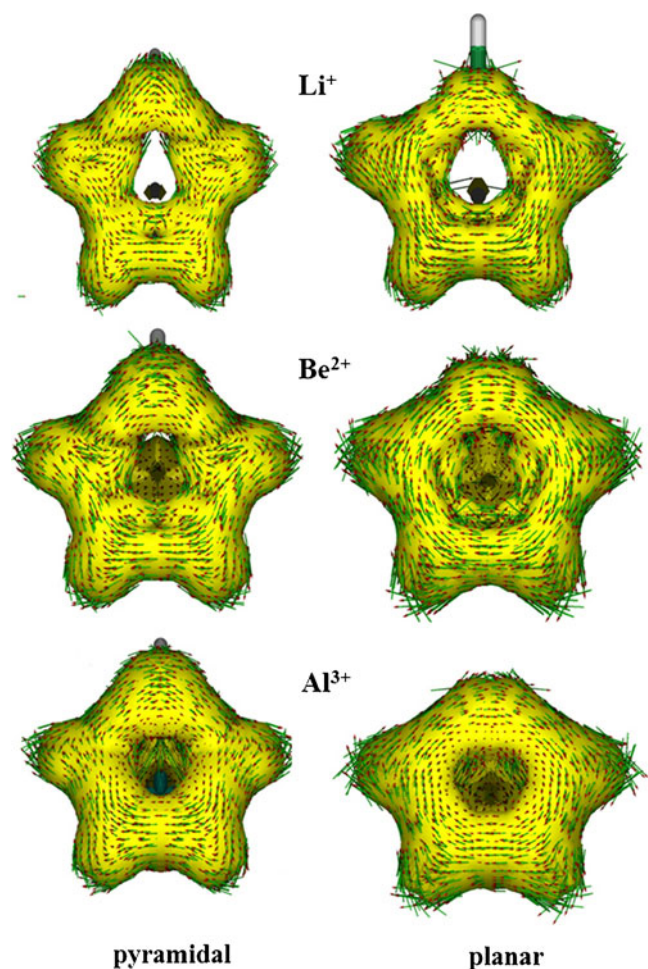
Fig. 3 ACID figures for the studied complexes

the effects of the local σ-bonding contributions are diminished. For this reason NICS values are also calculated at 1 Å away from the geometrical center of the ring in the opposite direction to cation (Fig. 2).

We should remember that the evaluation of the absolute aromaticity of a compound remains a controversial issue [34]. The main reason to this affirmation with regard to the magnetic properties as indication of aromatic character is the lack of a reference. For this reason, we have thought in the use of some system as reference value. Many previous works have analyzed the relationship between pyramidality and aromaticity, so we have chosen the planar structures corresponding to each system as reference structure. These planar structures are transition states (except for the phosphole-Al^{3+} planar structure which is a minimum at MP2 theory level).

The B3LYP and MP2 results are in agreement and show important facts. Firstly, the analysis of the difference in the values of magnetic properties for pyramidal and planar structure allows to indicate an enhancement in the aromaticity due to the complexation and this enhancement increase from Li^+ to Al^{3+}. On the other hand, we have to show the behavior for the Al^{3+} complex: in this case the pyramidal structure is even more aromatic than the planar structure according to χ and NICS values. Moreover, this case is a clear example of the importance of comparison in the use of NICS values. So, if we only pay attention to the NICS(0) value we can conclude a non-aromatic character because this is a positive value. Finally, the values for magnetic properties in the 1,2-BH_2-5-NH_2-phosphole are similar in the planar and pyramidal structure.

In order to carry out a deeper study of the aromaticity of the phosphole and these derivatives, the ACID method was employed. This is a relatively recently published method to investigate the delocalization and conjugation effects in molecules. It provides a powerful way to visualize the density of delocalized electrons and quantify conjugation effects. The ACID approach has several advantages: it is a scalar field which is invariant with respect to the relative orientation of the magnetic field and the molecule, it has the same symmetry as the wave function, and it can be plotted as an isosurface. In our group this method has been extensively employed in order to distinguish between pericyclic/pseudopericyclic and coarctate/pseudocoarctate reactivity, differentiation where the aromaticity has a crucial paper [51, 52].

Figure 3 presents the ACID isosurfaces for the complexes between phosphole and the studied cations at an isosurface value of 0.03 au. The ACID isosurfaces for the planar structures are also presented as comparison. Current density vectors are plotted onto the ACID isosurfaces. The first point to emphasize is that Fig. 3 shows a strong diatropic ring current for all these compounds, indicating their aromatic character. We have to emphasize the convergent results

obtained with NICS values and ACID figures: for the interaction with Li^+, the ring current is clearly stronger in the planar structure in agreement with the NICS values (NICS$_{planar}$ − NICS=7.72 for DFT structures). In the Be^{2+} case, the difference between the two structures is slighter than the difference in the NICS values (NICS$_{planar}$ − NICS= 2.07 for DFT structures). Finally, the ACID figures for complexes with Al^{3+} present in both cases strong ring currents, being even stronger for the pyramidal structure.

Conclusions

This work allows to us to indicate the increase of the aromatic character of the phosphole when complexes with Li^+, Be^{2+} and Al^{3+} are formed. This aromaticity increase from the Li^+ to the Al^{3+} case and is shown by the values of the magnetic properties, and the ACID figures. The agreement between conclusions obtained by magnetic properties and ACID method should be emphasized.

In order to avoid the problem of lack of reference of the magnetic properties as indicator of the aromaticity, a model was employed: the magnetic properties of the structures are compared with the magnetic properties of the planar structures.

Another important conclusion is that, for this type of complexes, the B3LYP method seems to be a good method. So, the structures at B3LYP and MP2 levels are quite similar and the qualitative results for inversion barriers are the same.

With the aim to examine the different effect of complexation or the use of π-acceptor groups at phosphorus, the 1,2-BH_2-5-NH_2-phosphole is studied. With a pyramidality very similar to Al^{3+}-phosphole, the inversion barrier is 2.38 kcal mol^{-1} lower and the aromaticity for pyramidal and planar structure is almost the same.

Finally, it is important to emphasize the strong effect of the Al^{3+} complexation. The knowledge of the behavior of the Al^{3+} with aromatic compounds is always interesting, according with the importance that the aluminum (III) cation has gained in the chemical and biochemical world after numerous negative aspects of it were reported.

References

1. McFail-Isom L, Shui X, Williams LD (1998) Biochemistry 37:17105–17111
2. Eichhorn GL (1981) Adv Inorg Biochem 3:1–46

3. Bock CW, Katz AK, Markham GD, Glusker JP (1999) J Am Chem Soc 121:7360–7372
4. Fraústo da Silva JJR, Williams RJP (1991) The biological chemistry of the elements. The inorganic chemistry of life. Clarendon, Oxford
5. Kim SK, Tarakeshwar P, Lee JY (2000) Chem Rev 100:4145–4186
6. Choi HS, Suh SB, Cho SJ, Kim KS (1998) Proc Natl Acad Sci USA 95:12094–12099
7. Ma JC, Dougherty DA (1997) Chem Rev 97:1303–1324
8. Ryzhov V, Dunbar RC, Cerda B, Wesdemiotis C (2000) J Am Soc Mass Spectrom 11:1037–1046
9. Amicangelo JC, Armentrout PB (2000) J Phys Chem A 104:11420–11432
10. Zhu W, Tan X, Shen J, Luo X, Cheng F, Mok PC, Ji R, Jiang H (2003) J Phys Chem A 107:2296–2303
11. Reddy AS, Sastry GN (2005) J Phys Chem A 109:8893–8903
12. Ruan C, Rodgers MT (2004) J Am Chem Soc 126:14600–14610
13. Dunbar RC (2000) J Phys Chem A 104:8067–8074
14. Costanzo F, Della Valle RG, Barone V (2005) J Phys Chem B 109:23016–23023
15. Choua S, Sidorenkova H, Berclaz T, Geoffroy M, Rosa P, Mezailles N, Ricard L, Mathey F, Le Floch P (2000) J Am Chem Soc 122:12227–12234
16. Deacon GB, Delbridge EE, Fallon GD, Jones C, Hibbs DE, Hursthouse MB, Skelton BW, White AH (2000) Organometallics 19:1713–1721
17. Gudat D, Bajorat V, Hap S, Nieger M, Schroder G (1999) Eur J Inorg Chem 1169–1174
18. Vijay D, Sastry GN (2006) J Phys Chem A 110:10148–10154
19. Braye EH, Hübel W (1959) Chem Ind London 1250
20. Leavitt FC, Manuel TA, Johnson F (1959) J Am Chem Soc 81:3163–3164
21. Nyulászi L (2001) Chem Rev 101:1229–1246
22. Quin LD, Bryson JG (1967) J Am Chem Soc 89:5984–5985
23. Mathey F (1969) C R Acad Sci Ser C 269:1066–1068
24. Mathey F (1988) Chem Rev 88:429–453
25. Hughes AN (1992) Phospholes and related Compounds. In: Handbook of Organophosphorus Chemistry. Dekker, New York
26. Quin LD, Hughes AN (1990) Cyclic Phosphines. In The Chemistry of Organophosphorus Chemistry. Wiley, Chichester
27. Coggon P, Engel JF, McPhail AT, Quin LD (1970) J Am Chem Soc 92:5779–5780
28. Nyulászi L (1995) J Phys Chem 99:586–591
29. Delaere D, Dransfeld A, Nguyen MT, Vanquickenborne LG (2000) J Org Chem 65:2631–2636
30. Quin LD, Keglevich Gy, Ionkin AS, Kalgutkar R, Szalontai G (1996) J Org Chem 61:7801–7807
31. Keglevich Gy, Böcskei Zs, Keserü Gy, Ujszászi K, Quin LD (1997) J Am Chem Soc 119:5095–5099
32. Nyulászi L, Soós L, Keglevich Gy (1998) J Organomet Chem 566:29–35
33. Nyulászi L (1996) J Phys Chem 100:6194–6198
34. See special issue 5 entirely dedicated to aromaticity (2001) Chem Rev 101:115–1566
35. Zimmermann HE (1971) Acc Chem Res 4:272–280
36. Herges R, Jiao H, Schleyer PvR (1994) Angew Chem Int Ed 33:1376–1378
37. Jiao H, Schleyer PvR (1998) J Phys Org Chem 11:655–662
38. Schleyer PvR, Maerker C, Dransfeld A, Jiao H, Hommes NJRvE (1996) J Am Chem Soc 118:6317–6318
39. Herges R, Geuenich D (2001) J Phys Chem A 105:3214–3220
40. Herges R, Papafilippopoulos A (2001) Angew Chem Int Ed 40:4671–4674
41. Kimball DB, Weakley TJR, Herges R, Haley MM (2002) J Am Chem Soc 124:13463–13473
42. Becke AD (1993) J Chem Phys 98:5648–5652
43. Lee C, Yang W, Parr RG (1998) J Phys Rev B 37:785–789
44. Wolinski K, Hilton JF, Pulay P (1990) J Am Chem Soc 112:8251–8260
45. Keith TA, Bader RFW (1993) Chem Phys Lett 210:223–231
46. Keith TA, Bader RFW (1992) Chem Phys Lett 194:1–8
47. Cheeseman JR, Frisch MJ, Trucks GW, Keith TA (1996) J Chem Phys 104:5497–5509
48. 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, 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, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98. Gaussian Inc, Pittsburgh, PA
49. 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, 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, Wallingford, CT
50. Schleyer PvR, Manoharan M, Wang Z-X, Kiran B, Jiao H, Puchta R, Hommes NJRvE (2001) Org Lett 3:2465–2468
51. Peña-Gallego A, Rodríguez-Otero J, Cabaleiro-Lago EM (2004) J Org Chem 69:7013–7017
52. Peña-Gallego A, Rodríguez-Otero J, Cabaleiro-Lago EM (2007) J Phys Chem A 111:2935–2940

ORIGINAL PAPER

Noncovalent and covalent functionalization of a (5, 0) single-walled carbon nanotube with alanine and alanine radicals

Muthusivarajan Rajarajeswari · Kombiah Iyakutti · Yoshiyuki Kawazoe

Abstract We have systematically investigated the non-covalent and covalent adsorption of alanine and alanine radicals, respectively, onto a (5, 0) single-walled carbon nanotube using first-principles calculation. It was found that XH···π (X=N, O, C) interactions play a crucial role in the non-ovalent adsorption and that the functional group close to the carbon nanotube exhibits a significant influence on the binding strength. Noncovalent functionalization of the carbon nanotube with alanine enhances the conductivity of the metallic (5, 0) nanotube. In the covalent adsorption of each alanine radical onto a carbon nanotube, the binding energy depends on the adsorption site on CNT and the electronegative atom that binds with the CNT. The strongest complex is formed when the alanine radical interacts with a (5, 0) carbon nanotube through the amine group. In some cases, the covalent interaction of the alanine radical introduces a half-filled band at the Fermi level due to the local sp^3 hybridization, which modifies the conductivity of the tube.

Keywords Adsorption · Carbon nanotubes · Amino acids · Density functional theory

M. Rajarajeswari · K. Iyakutti (✉)
School of Physics, Madurai Kamaraj University,
Madurai,
Tamil Nadu 625 021, India
e-mail: iyakutti@yahoo.co.in

Y. Kawazoe
Institute for Materials Research, Tohoku University,
Sendai 980–8577, Japan

Introduction

Hybrids of hollow cylindrical single-walled carbon nanotubes (SWCNT) fused with biomolecules are finding applications in a wide range of fields, such as molecular-level electronics [1, 2], biochemical sensors [3–6], drug delivery [7], gene delivery [8, 9], as well as in therapeutic applications [10]. The integration of biomolecules with CNT can be achieved by either a covalent or a noncovalent approach. A broad variety of organic and biomolecules such as proteins, peptides, and nucleic acid bases are used to functionalize the CNT. Metalloproteins and enzymes are adsorbed onto the CNTs in aqueous solution, and AFM studies confirm this as a physical immobilization of the adsorbed species [3]. The noncovalent binding properties of protein and DNA are utilized in intercellular transport via CNT [11]. One of the biggest challenges when attempting to make use of carbon nanotubes for practical applications is their poor solubility. Aromatic amino acids in the peptide sequence facilitate the effective isolation of individual CNTs [12], and the binding of peptides is sensitive to the chirality of the CNT [13]. The spontaneous adsorption of the peptides NB1, B1, and B3 onto the CNT side wall has been investigated using quantum mechanical method [14]. The aromatic amino acids histidine and tryptophan play a key role in peptide adsorption and enhance the intermolecular interaction with CNT. The mechanism of interaction of a peptide with a CNT depends on the nature of the amino acid that is in close contact with the CNT [15].

Proteins (also known as polypeptides) are organic compounds consisting of amino acids that are arranged in a linear chain and connected by peptide bonds. Thus, the

individual amino acids in a protein influence the binding of the protein with the CNT [12, 14, 15]. Investigating the interactions of individual amino acids with the CNT will lead to a better understanding of the biofunctionalization of CNTs. With modern methods, we can tether the desired amino acid to the CNT for particular applications. Several studies of the noncovalent adsorption of amino acids onto carbon nanostructures have been performed using various computational techniques [16–21]. The adsorptions of the amino acids glycine, histidine, phenylalanine and cysteine onto a (3, 3) CNT [16] and a (10, 0) CNT [17] have been studied in detail. Sun et al. [18] investigated the interaction of glycine/glycine radicals with intrinsic/boron-doped (8, 0) SWCNTs. Glycine was physisorbed onto intrinsic CNT and chemisorbed onto boron-doped CNT due to the presence of the electron-poor boron atom. The interactions of all 20 amino acids with a capped (5, 5) CNT [19] and (9, 9) CNT with one Li atom inside the tube [20] have been reported. The adsorption of aromatic amino acids onto graphene and onto a (5, 5) CNT was exclusively investigated by Rajesh et al. [21]. The binding strengths of amino acids with a CNT and a graphene substrate were found to be directly correlated with the electronic polarizability.

In the covalent method, an amino acid is directly bound to a CNT through a chemical bond or through a linker molecule. Organic molecules functionalized with CNTs act as linkers in order to couple an amino acid or peptide with CNT [22–25]. Hydrophilic poly-L-lysine was covalently linked with an CNT through a carboxylic acid group on the CNT side wall [23]. The addition of peroxidase to a CNT and poly-L-lysine complex can be used in H_2O_2 biosensing. Through a one-step substitution, amino-functionalized carbon nanotubes can be formed with the aid of fluorinated nanotubes, and this increases the solubility of CNTs in polar solvents [24]. The covalent adsorption of alanine radicals onto pristine and functionalized (5, 5) CNTs has been reported by Carneiro et al. [25].

In the above reports, the importance of the functionalization of carbon nanotubes with amino acids is demonstrated. In some previous studies [19, 20], binding energies for the noncovalent adsorption of amino acids onto carbon nanotubes were presented, but details on the adsorption mechanism, the nature of the interaction and electronic structure analysis were not discussed. Leon et al. [19] reported that the amino acids arganine, cysteine and alanine exhibited the strongest interactions with (5, 5) CNTs. Interesting results were also presented by Carneiro et al. [25] regarding the covalent adsorption of alanine onto (5, 5) CNTs. Among the twenty available amino acids, after glycine, alanine is the simplest and smallest amino acid. In previous investigations [16–21, 25], (3,3), (5, 5), (9, 9), (8,0), and (10, 0) CNTs were considered for amino acid functionalization. Here, for the first time, a (5, 0) CNT was

considered for amino acid functionalization. The (5, 0) CNT is a small-diameter zigzag metallic nanotube, so the binding energy involved is expected to be large. Also, it is computationally less expensive to investigate the interaction between a (5, 0) CNT and alanine. A systematic study of the interaction between this CNT and the amino acid alanine will shed light on the functionalization of CNTs with amino acids, and will also indicate possible applications of such CNTs. In this report, we present a detailed description of the noncovalent and covalent functionalization of a (5, 0) single-walled carbon nanotube with the amino acid alanine and alanine radicals, respectively, using density functional theory.

Computational details

Density functional calculations were performed using plane-wave basis sets as implemented in the Vienna ab initio simulation package (VASP) [26, 27]. The projector augmented-wave method (PAW) [28, 29] was used to describe the interaction between ions and electrons. It is known that for noncovalent systems, the local density approximation (LDA) [30] overestimates the binding energy, while the generalized gradient approximation (GGA) [31] underestimates it. The calculated interlayer distance for the adsorption of adenine onto a graphene sheet within the LDA is almost equal to its experimental value [32]. In the study of the noncovalent adsorption of aromatic molecules onto a nanotube, the binding energies calculated with LDA and hybrid functionals were found to be nearly the same [33]. Thus, within LDA, the interaction energies of noncovalent systems are well described, and LDA can provide a good description of dispersive interactions within certain limits as compared to the GGA [34, 35]. Also, the reported binding energies for the covalent adsorption of alanine radicals onto CNTs within LDA and GGA are qualitatively similar [25]. Hence, we have adopted the LDA approach for our study. A gamma-centered 1×1×5 Monkhorst–Pack grid [36] is used for k-point sampling of the Brillouin zone. For electronic band structure calculations, 81 irreducible k-points are considered in the z-axis direction of CNT. The cutoff energy for the plane-wave basis set of the valence electrons was set to be 500 eV throughout the calculations. The residual minimization–direct inversion in the iterative subspace method (RMM-DIIS) was used for wavefunction optimization. Geometry optimization was carried out until the magnitude of the forces acting on all of the atoms was smaller than 10^{-5} eV/Å.

In this paper, we report an investigation of the noncovalent and covalent adsorption of the amino acid alanine onto a (5, 0) CNT. Alanine is a nonpolar hydrophobic molecule, and the α carbon of alanine is optically active.

As is generally the case for proteins, only the L-isomer form of alanine is found naturally and thus we have only considered the L-isomer in our study. This is the simplest amino acid, with one amine, carboxyl and methyl group. Initially, we optimized the CNT and alanine separately. In the hexagon ring of the CNT, two sites are available for noncovalent alanine adsorption, namely the hollow hexagonal center (Fig. 1a) and the bridge site along the C–C bond (Fig. 1b). Alanine is allowed to interact with the CNT through the functional groups CH_3, NH_2, COOH and OH in different orientations. The potential energy surface for the adsorption of alanine onto the CNT surface is traced out by two steps: the first involves the alanine moving from the surface of the CNT in a 0.5 Å step, and then shifts into different orientations above the hexagonal and bridge sites. As a result, we get 12 different CNT+alanine conformers. The (5, 0) CNT+alanine complex with the respective functional group (for example CH_3) near the CNT is denoted $C_{50}ACH_3$. Throughout this paper, this notation is used to represent the conformers considered in this study. To investigate the covalent interaction of alanine with the CNT, we created four different alanine radicals: (1) by removing the H of the NH_2; (2) by removing the H_2 of the NH_2; (3) by removing the H of the COOH, and; (4) by removing the OH of the COOH. The alanine radicals were allowed to interact with the C–C bond of the CNT at two different sites, as shown in Fig. 1c. The supercell approach was followed with a lateral separation of 24 Å in the x- and y-directions to avoid interactions with the periodic images. The supercell had two unit cells of (5, 0) CNT with a length of 7.91 Å.

Results and discussion

Noncovalent adsorption of alanine onto (5, 0) CNT

Structures and binding energies

After the geometrical optimization, we considered 12 different conformers in which the alanine is bound to the CNT through one of the functional groups (Fig. 2). Equilibrium distances between the CNT and alanine and the binding energy of alanine are presented in Table 1. The effective interaction energy of the noncovalent complex is
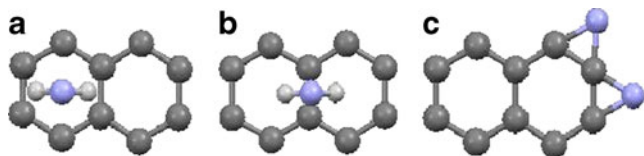
the sum of the electrostatic, induction, charge transfer, dispersion and repulsive interactions. Electrostatic and van der Waals interactions play a crucial role in the noncovalent adsorption. Functional groups (CH_3, NH_2, COOH, and OH) of alanine can interact with the CNT π-electron cloud through $XH\cdots\pi$ (X=C, N and O) interactions. In addition the hydrophobic nature of the CNT also enhances interaction. Though the adsorption of alanine onto the CNT is physisorption, there can be a partial electron transfer between the species. For all twelve conformers, we calculated the charge transfer between the CNT and alanine using Bader analysis [37]. Charge transfer is one of the key factors that facilitates the adsorption of adsorbate on the adsorbent. Electron-rich atoms like nitrogen and oxygen play an important role in electron donor–acceptor interactions. The calculated charge transfer is given in Table 1. In some cases, charge is transferred from CNT to alanine (shown as positive), while in others charge is transferred from alanine to CNT (shown as negative). In most cases the charge transfer is negligible.

Adsorption of alanine through the methyl group gave two stable conformers, $C_{50}ACH_3$-I (Fig. 2a) and $C_{50}ACH_3$-II (Fig. 2b). The methyl group of alanine is positioned on the CNT surface at two sites: either just above the hollow hexagonal center (Fig. 2a) or on the bridge site (Fig. 2b). The equilibrium distances between alanine and CNT for the two conformers are nearly the same, but different binding strengths with CNT are observed. In the conformer $C_{50}ACH_3$-I, the distance between the C (CNT) and H atom is around 2.5 Å, which places the hydrogen atoms of CH_3 in direct contact with the π-electron cloud of the CNT. Therefore, the attraction between the CNT and alanine arises from the $CH\cdots\pi$ interaction, and the conformer is stabilized with a large binding energy. In the conformer $C_{50}ACH_3$-II, the functional group CH_3 is present above the bridge. The distance between the hydrogen (CH_3) and carbon atom (CNT) is around 2.9 Å, and this reduces the stabilization energy of the complex. Leon et al. [19] reported that the adsorption of alanine onto a (5, 5) CNT led to a stable complex with a binding energy of 5.4 kcal mol^{-1} (i.e., 0.23 eV). In that case, the CH_3 group of alanine is close to the surface of the CNT (2.85 Å from it). In a recent paper, Chang and Jalbout [20] reported the effect of alkali metal encapsulation inside a (9, 9) armchair nanotube on its interaction with an externally adsorbed amino acid, and stated that the binding energy of alanine was 69.5 kJ mol^{-1} (i.e., 0.72 eV). In our study, we found that alanine was adsorbed at a distance of 3.6 Å above the hollow hexagonal site ($C_{50}ACH_3$-I) with a binding energy of 0.76 eV, and at a distance of 3.4 Å from the bridge site ($C_{50}ACH_3$-II) with a binding energy of 0.11 eV. Our calculated binding energies and equilibrium distances (Table 1) for the noncovalent adsorption of alanine on a

Fig. 1 a–c Possible adsorption sites on the CNT hexagon. **a** Hexagonal hollow center. **b** Bridge site. **c** Cross bridge and bridge site
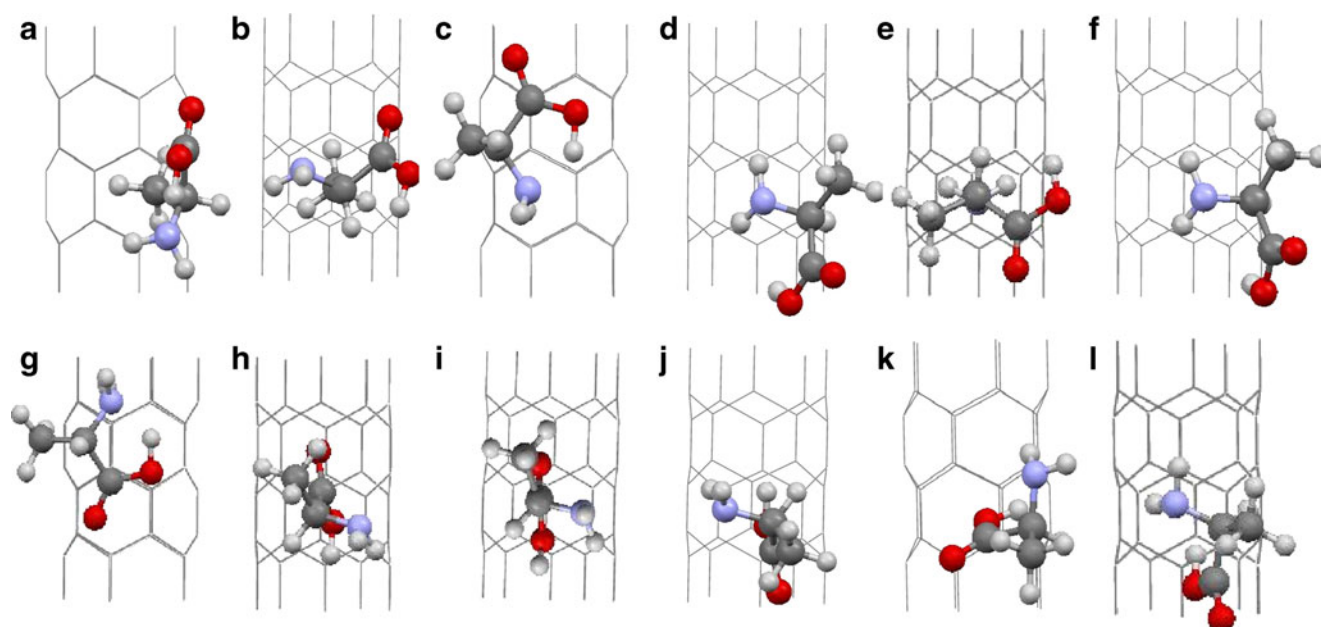
Fig. 2 a–l Noncovalent adsorption of alanine on (5, 0) CNT. a $C_{50}ACH_3$-I. b $C_{50}ACH_3$-II. c $C_{50}ANH_2$-I. d $C_{50}ANH_2$-II. e $C_{50}ANH_2$-III. f $C_{50}ANH_2$-IV. g $C_{50}ACOOH$-I. h $C_{50}ACOOH$-II. i $C_{50}ACOOH$-III. j $C_{50}AOH$-I. k $C_{50}AOH$-II. l $C_{50}AOH$-III

(5, 0) CNT are in agreement with previously reported values [19, 20].

Then we considered the adsorption of alanine onto a CNT through the amine group. The lone pair of electrons of the nitrogen are attracted to the π-electron cloud, leading to charge transfer from alanine to the CNT. The amine group of alanine can interact with the π-electron cloud of the CNT through the formation of a hydrogen bond via NH⋯π

Table 1 Equilibrium distances, charge transfer values and binding energies for the noncovalent adsorption of alanine on (5, 0) CNT

Conformers	Distance[a] $(d, Å)$	Charge (q, e)[b]	Binding energy (E_b, eV)
$C_{50}ACH_3$-I	3.6	0.006	0.76
$C_{50}ACH_3$-II	3.4	0.004	0.11
$C_{50}ANH_2$-I	3.6	−0.060	0.95
$C_{50}ANH_2$-II	3.2	−0.122	0.50
$C_{50}ANH_2$-III	3.0	−0.015	0.17
$C_{50}ANH_2$-IV	3.0	−0.115	0.47
$C_{50}ACOOH$-I	3.2	0.015	0.92
$C_{50}ACOOH$-II	3.2	−0.024	0.21
$C_{50}ACOOH$-III	2.8	−0.008	0.15
$C_{50}AOH$-I	3.1	0.005	0.20
$C_{50}AOH$-II	3.1	0.007	0.08
$C_{50}AOH$-III	2.9	0.005	0.26

[a] Distance XH⋯C, X=N, O, C of alanine

[b] CNT to alanine:positive; alanine to CNT: negative

interactions. In this case, the CNT+alanine complex is stabilized by the electrostatic, charge transfer and dispersion energy terms. All four conformers exhibit different binding strengths depending on the orientation of the adsorbate and the charge transfer. The conformer $C_{50}ANH_2$-I (Fig. 2c) exhibits greater binding energy than all of the other noncovalent systems discussed in this paper. The NH_2 of $C_{50}ANH_2$-I is positioned at the center of the hollow hexagonal site at a distance of 3.6 Å. One of the hydrogen atoms of the NH_2 faces this hexagonal center and forms a hydrogen bond with the carbon atom of the CNT. Even though charge transfer to the CNT from the alanine is rather limited (0.06e), the NH⋯π interaction establishes the complex. The electrostatic repulsive interaction between the alanine and CNT is diminished due to the large separation between them (3.6 Å). Thus, the overall interaction becomes attractive and this system is more stable than the other conformers. Among the four conformers, $C_{50}ANH_2$-III (Fig. 2e) is stabilized with a small binding energy; again, the charge transfer from alanine to the CNT is negligible in this case (0.015e). This is mainly because of the orientation of the NH_2 on the CNT surface. The NH_2 is positioned horizontally above the bridge C–C site, and the distance between the hydrogen and the carbon atom of the CNT is 2.7 Å. The angle between the NH and C does not favor the charge transfer and interaction with the π-electron cloud. Thus, the binding energy is lower for this conformer than the others.

Among the twelve conformers, the conformers $C_{50}ANH_2$-II (Fig. 2d) and $C_{50}ANH_2$-IV (Fig. 2f) show the largest transfer of charge (0.1e) from alanine to the

CNT. In both of these conformers, the NH_2 group is oriented vertically, and the carbon atom of the CNT makes direct contact with the hydrogen atoms of the NH_2. However, the binding energy of $C_{50}ANH_2$-II is higher than that of $C_{50}ANH_2$-IV. In $C_{50}ANH_2$-II, the hydrogen atoms of amine group are very close (2.4 Å) to the carbon atoms of the CNT. Two hydrogen bonds are formed with the nearest carbon atoms in the CNT hexagon. The dipole-induced-dipole interaction stabilizes the conformer with a high binding energy. In the case of $C_{50}ANH_2$-IV, the orientation of amine group and the distance between the hydrogen and carbon atoms (2.7 Å) does not favor the interaction, so the binding is weak.

The interaction of alanine with CNT through the COOH group gives rise to three stable conformers. In the conformer $C_{50}ACOOH$-I (Fig. 2g), O and OH are positioned at the hollow hexagonal site at 3.2 Å from the surface of the CNT, and the OH group is shifted to the upper hollow hexagonal site. There is no hydrogen bond formation, but 0.015e of charge is transferred from the CNT to alanine. An attractive interaction arises between the π-electron cloud and the lone pair of electrons of the oxygen atom, and this stabilizes the complex. With a binding energy of 0.92 eV, this is the second most stable conformer among those studied. In the conformer $C_{50}ACOOH$-II (Fig. 2h), the functional group is oriented vertically above the hollow hexagonal site and exhibits a binding energy of 0.21 eV, which is very low compared to the previous case. In the third case, COOH is positioned vertically above the bridge site (Fig. 2i) at a distance of 2.8 Å from the CNT surface, with a binding energy of 0.15 eV. Here, the COOH group is perpendicular to the CNT surface and the hydrogen atom of the COOH is projected away from the surface of the CNT. Thus, the unfavorable orientation of COOH minimizes the interaction with the π-electron cloud of the CNT, resulting in reduced binding energy.

In the conformer $C_{50}AOH$-I (Fig. 2j), the OH group is positioned vertically above the hollow hexagonal site, with a binding energy of 0.20 eV. In the case of $C_{50}AOH$-II (Fig. 2k), the OH is above the bridge site, in a tilted horizontal orientation that has the lowest binding energy of 0.08 eV. In both conformers, the OH group is 3.1 Å from the surface of the CNT, and the C–H distance is around 2.7 Å.

This implies a weak electrostatic interaction between the alanine and the CNT. The OH of the conformer $C_{50}AOH$-III (Fig. 2l) is shifted down to the next hexagonal site, and one hydrogen bond is formed with the carbon atom of the CNT, with a bond length of 2.2 Å. Thus, the electrostatic interaction between the OH group and the CNT surface is manifested through the dipole–induced-dipole interaction.

Charge density analysis

Charge density analysis was performed to get a better understanding of the interaction mechanism for the non-covalent adsorption of alanine onto the (5, 0) CNT. The molecular orbitals of some selective conformers $C_{50}ACH_3$-I and $C_{50}ANH_2$-I are shown in Fig. 3 [38]. In the highest occupied molecular orbital (HOMO) of the conformer $C_{50}ACH_3$-I (Fig. 3a), the charge is distributed over the alanine homogeneously. The accumulated charge on the alanine leads to an electrostatic interaction with the CNT, and subsequently to one of the most stable conformers. In the HOMO of $C_{50}ANH_2$-I (Fig. 3c), the charge is concentrated on the nitrogen and oxygen atoms. Here, however, alanine interacts through the amine group; the carboxyl group is also close to the CNT in the optimized structure. This enhances the charge transfer from both the nitrogen and the oxygen atoms, and thus the alanine possesses a maximum charge of 0.1e. In the lowest unoccupied molecular orbital (LUMO) of $C_{50}ANH_2$-I (Fig. 3d), charge is distributed uniformly on the CNT.

Electronic structures

The (5, 0) CNT is a narrow nanotube with a diameter of 3.96 Å. Theoretically, the (5, 0) CNT is predicted to be a semiconductor without σ–π hybridization. However, computational calculations show that it undergoes a semiconductor to metal transition due to the σ–π hybridization [39, 40]. Our band structure calculation also showed that the pristine (5, 0) CNT is metallic, with bands crossing the Fermi energy (Fig. 4a). Adsorption of alanine onto the CNT modifies the bands near the Fermi energy, and in all cases the Fermi energy is blueshifted. Figure 4b–e shows the band structures of some selected conformers in which we
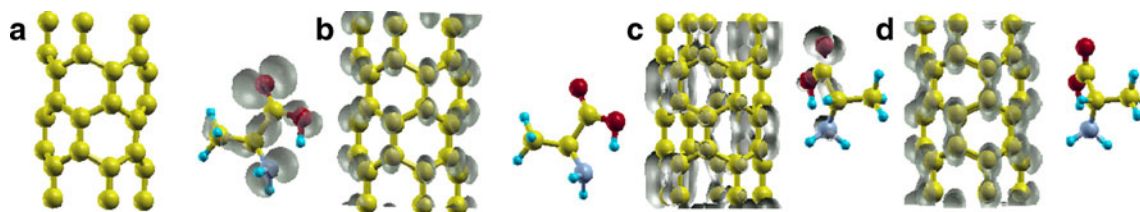
Fig. 3 **a** Highest occupied molecular orbital (HOMO) of $C_{50}ACH_3$-I ($E = -0.12$ eV). **b** Lowest unoccupied molecular orbital (LUMO) of $C_{50}ACH_3$-I ($E = 0.38$ eV). **c** HOMO of $C_{50}ANH_2$-II ($E = -0.23$ eV. **d** LUMO of $C_{50}ANH_2$-II ($E = 0.25$ eV)
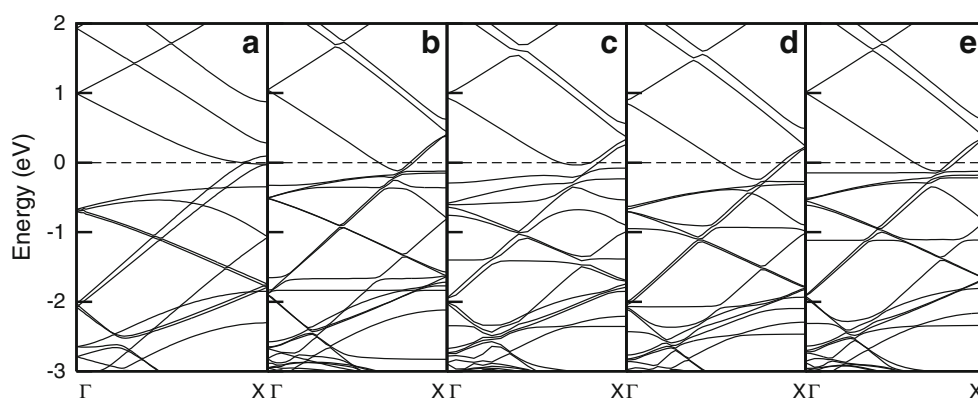
observed band rearrangements near the Fermi energy. In $C_{50}ACH_3$-I (Fig. 4b), the adsorption of alanine modified the bands of the (5, 0) CNT near the Fermi energy, and this is clearly reflected in the density of states (DOS) plot (Fig. 5b). A new band appeared just below the Fermi energy, and a flat band arose at −1.9 eV due to the nitrogen atom of the alanine. A peak near the Fermi energy in the DOS plot of the pristine (5, 0) CNT was shifted towards the Fermi energy in the DOS plot of $C_{50}ACH_3$-I, which caused a slight increase in the density of states. Also, there was a new Van Hove singularity at −1.9 eV, corresponding to the flat band in the band structure.

In the case of $C_{50}ANH_2$-II, the amine group of the alanine is positioned close the CNT, and additional bands appear near the Fermi energy due to the nitrogen atom (Fig. 4c). A band originating from the valance band region at −0.6 eV (Γ point) passes into the conduction bands, but the band originating from the conduction band region is just about at the Fermi energy, as in the pristine (5, 0) CNT. This increases the DOS at the Fermi energy (Fig. 5c) and enhances the conductivity. A new band is present just below the Fermi energy, and this appears as a shoulder of the peak in the DOS due to the nitrogen atom. As in the $C_{50}ACH_3$-I conformer, a new Van Hove singularity appeared at around −1.9 eV. Here, both the nitrogen and oxygen atoms contribute to the Van Hove singularity, because in the optimized structure of $C_{50}ANH_2$-II, the OH group is close to the surface of the CNT in addition to the amine group. The projected density of states (PDOS) confirms that the increase in the density of states at the Fermi energy is due to the contribution of the nitrogen's p orbitals.

In $C_{50}ACOOH$-I (Fig. 4d), the first conduction band crosses the Fermi energy and dips down into the valence band. No additional bands appear near the Fermi energy. Therefore, in the DOS plot (Fig. 5d), only one peak is present at the Fermi energy. One new band appears at around −1 eV due to the nitrogen atoms. The band structure of $C_{50}AOH$-III (Fig. 4e) shows a flat band just below the Fermi energy, and some additional bands around −1 eV and below −2.5 eV. In the DOS plot (Fig. 5e), peaks are prominent and sharp. The

DOS at the Fermi energy is higher than in the other cases. Here, the second peak in the lower-energy region arises due to the nitrogen and oxygen atoms. In the PDOS, the peaks in the region of −4 to −6 eV appear due to the oxygen atoms, and they are prominent compared to those of the other conformers.
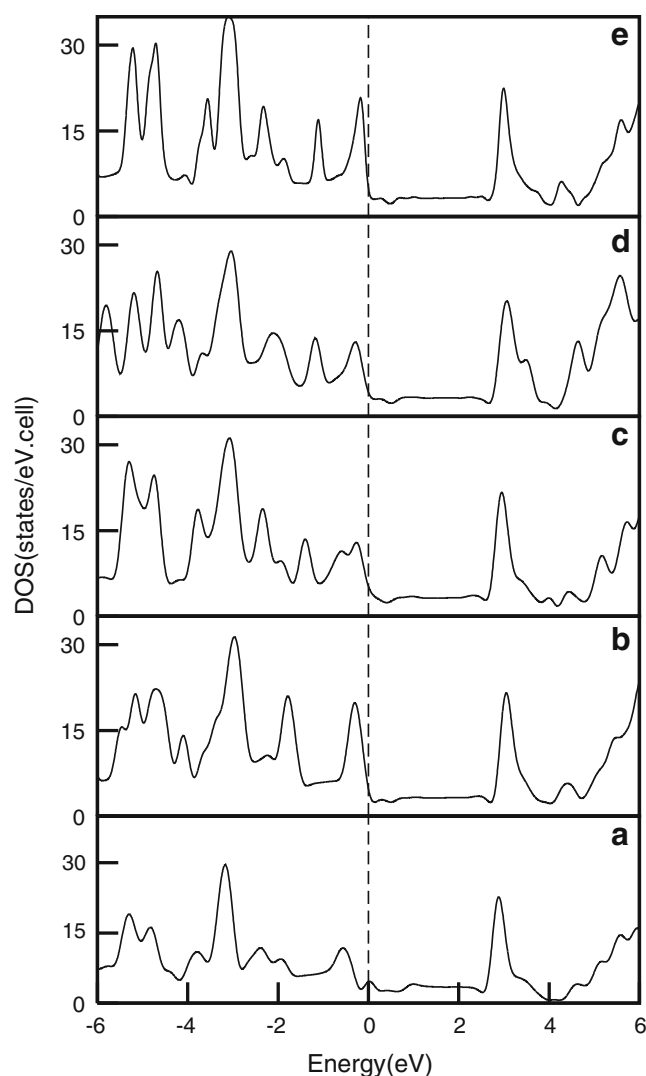
Fig. 5 a–e Density of states plots for **a** pristine (5, 0) CNT, **b** $C_{50}ACH_3$-I, **c** $C_{50}ANH_2$-II, **d** $C_{50}ACOOH$-I, **e** $C_{50}AOH$-III
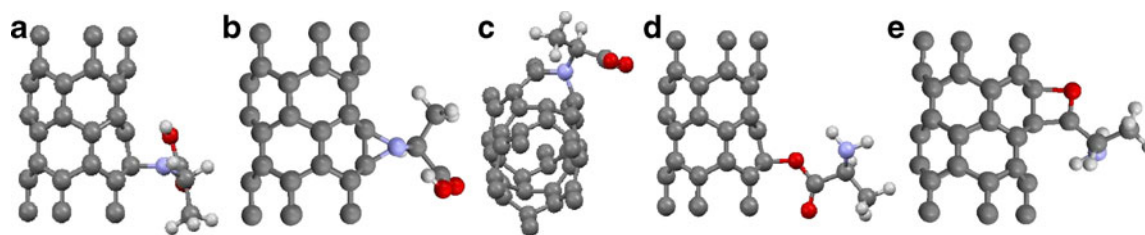
Fig. 6 a–e Covalent adsorption of alanine radicals onto the (5, 0) CNT. **a** C_{50}ANH-I. **b** C_{50}AN-I. **c** C_{50}AN-II. **d** C_{50}ACOO-I. **e** C_{50}ACO-I

Covalent adsorption of alanine onto (5, 0) CNT

Structures and binding energies

In this section, we report the interactions of four different alanine radicals with the pristine (5, 0) CNT. After optimization, five stable conformers were identified (Fig. 6). Alanine radicals are allowed to interact with the C–C bonds in the CNT. Bader charge density analysis [37] was performed to determine the charge transfer. Equilibrium distances, charge transfer values and binding energies are presented in Table 2. The conformer C_{50}ANH-I (Fig. 6a) is obtained by removing the H atom from the amine group, and the NH is then allowed to interact directly with the carbon atoms at the bridge site. After structural optimization, the nitrogen atom of the amine group binds to one of the carbon atoms at the C–C bridge site, and the carbon atom protrudes from its original position in the direction of alanine (Fig. 6a). The C–N bond distance was found to be 1.45 Å, and the binding energy was 3.89 eV. The three other C–C bonds arising from that particular carbon atom were also elongated. Thus, the covalent attachment of the alanine radical to the CNT modifies the symmetry of the CNT and the carbon atom at the adsorption site becomes four coordinated. This induces sp^3 hybridization locally.

The conformers C_{50}AN-I (Fig. 6b) and C_{50}AN-II (Fig. 6c) are formed by removing H_2 from the amine group

of alanine. The nitrogen atom in the conformer C_{50}AN-I is bound covalently to the carbon atoms at the C–C bridge site. Due to these two elongated C–N bonds, the carbon atoms at the C–C bridge site are four coordinated. This conformer has a larger binding energy (6.40 eV) than the other conformers considered in this study. In the conformer C_{50}AN-II (Fig. 6c), the nitrogen atom of alanine has broken one of the C–C bonds in the CNT hexagon and bonded to the two carbon atoms. A new seven-membered ring with sp^2 bonding is thus created at the adsorption site. A maximum charge of 0.74e is transferred to the alanine radical from the CNT. Among the five structures investigated in our study of the covalent adsorption of alanine radicals onto the CNT, the conformer C_{50}AN-II alone retains sp^2 like hybridization. The same structural configurations (C_{50}AN-I and II) are reported for alanine radical adsorption onto the (5, 5) CNT [25]. The binding energy of structure III [25] (like C_{50}AN-II) is found to be 6.8 eV within the LDA, with a C–N distance of 1.41 Å. Compared to the value reported above, the binding energy (5.67 eV) we obtained was lower, due to the zigzag structure of our CNT. Similar to the conformers C_{50}ANH-I and C_{50}AN-I, the carbon atom at the adsorption site of the conformer C_{50}ACOO-I (Fig. 6d) protrudes slightly and is four coordinated with the oxygen atom of the alanine radical. The binding energy of the alanine radical is found to be 0.99 eV, which is lower than in the other cases. In the conformer C_{50}ACO-I (Fig. 6e), the OH group of alanine is removed. In the relaxed structure, the carbon and oxygen atoms of the alanine radical form two elongated bonds with the carbon atoms at the C–C bridge site, with bond lengths of 1.6 Å (C–O) and 1.47 Å (C–C). Thus, the carbon atoms at the bridge site are four coordinated, which induces local sp^3 hybridization.

Charge density analysis

Alanine radical adsorption onto CNT gives five different structures. The bond lengths in the conformers confirm the covalent nature of the adsorption. Further analysis of the charge density profile will afford a better understanding of the bonding between alanine and CNT. The highest occupied molecular orbitals (HOMOs) and the lowest unoccupied molecular orbitals (LUMOs) of selected con-

Table 2 Equilibrium distances, charge transfer values and binding energies for the covalent adsorption of alanine radicals onto (5, 0) CNT

Conformers	Distance $(d, \text{Å})^a$	Charge $(q, e)^b$	Binding energy (E_b, eV)
C_{50}ANH-I	1.45	0.23	3.89
C_{50}AN-I	1.45	0.54	6.40
	1.46		
C_{50}AN-II	1.38	0.74	5.67
	1.40		
C_{50}ACOO-I	1.47	0.56	0.99
C_{50}ACO-I	1.60	−0.03	3.50
	1.47		

^{a} Distance XH···C, X=N, O, C of alanine

^{b} CNT to alanine: positive; alanine to CNT: negative

Fig. 7 a–h a Highest occupied molecular orbital (HOMO) of $C_{50}ANH$-I ($E = -0.42$ eV). **b** Lowest unoccupied molecular orbital (LUMO) of $C_{50}ANH$-I ($E = 0.12$ eV). **c** HOMO of $C_{50}AN$-I ($E = -0.31$ eV. **d** LUMO of $C_{50}AN$-I ($E = 0.17$ eV). **e** HOMO of $C_{50}AN$-II ($E = -0.25$ eV). **f** LUMO of $C_{50}AN$-II ($E = 0.12$ eV). **g** HOMO of $C_{50}ACO$-I ($E = -0.29$ eV). **h** LUMO of $C_{50}ACO$-I ($E = 0.53$ eV)

formers are shown in Fig. 7. In the HOMO of the conformer $C_{50}ANH$-I (Fig. 7a), charge is accumulated in the CNT and at the nitrogen atom which is in contact with the CNT. However, the carbon atom that is in direct contact with the alanine does not possess any charge. This validates the excess charge found with alanine using Bader analysis. The properties of the HOMO and LUMO of the conformer $C_{50}AN$-I (Fig. 7c, d) justify the large binding energy of the alanine radical and the charge transfer from the CNT to alanine. Here, the nitrogen of the alanine radical forms two covalent bonds with the carbon atoms in the C–C bond. The HOMO shows the charge distribution on the α carbon, the carboxyl group and along one C–N bond. In the LUMO, the charge is distributed over the carbon atoms of the CNT alone. In the HOMO of the conformer $C_{50}AN$-II (Fig. 7e), charge is distributed over the nitrogen atom, which drags 0.74e from the CNT. In the LUMO (Fig. 7f), charge is distributed homogeneously. The HOMO of

$C_{50}ACO$-I (Fig. 7g) shows the accumulation of charge density on the alanine and carbon atoms near the adsorption site. In the LUMO (Fig. 7h), charge is distributed only on the alanine.

Electronic structures

The band structures for the adsorption of alanine radicals onto the CNT are presented in Fig. 8. In the band structure for the conformer $C_{50}ANH$-I (Fig. 8b), degeneracy is removed in some of the bands due to the adsorption of the alanine radical, and the Fermi energy is blueshifted. A half-filled band emerges out of the conduction band region at 0.2 eV due to the sp^3 defect induced by the alanine radical adsorption [41]. The density of states at the Fermi energy is slightly lower, and some new peaks appear in the lower and higher energy regions. The PDOS shows that the peaks formed in the lower-energy region are due to the
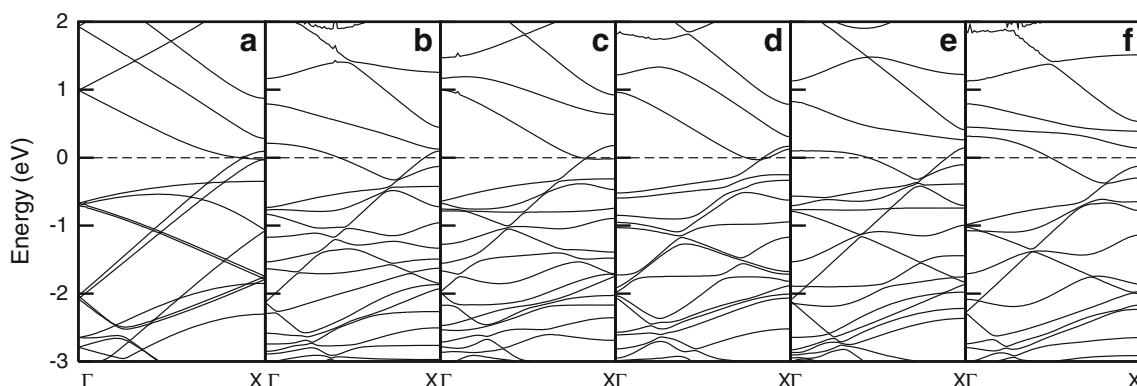
Fig. 8 a–f Electronic band structures for **a** pristine (5, 0) CNT, **b** $C_{50}ANH$-I, **c** $C_{50}AN$-I, **d** $C_{50}AN$-II, **e** $C_{50}ACOO$-I, **f)** $C_{50}ACO$-I

carbon, oxygen and hydrogen atoms, while the peaks in the higher energy region are due to the nitrogen atoms.

In the band structure of the conformer $C_{50}AN$-I (Fig. 8c), bands cross the Fermi energy, just as in the pristine CNT. This indicates that the CNT retains its metallic nature. Compared to the pristine CNT, the DOS (Fig. 9c) at the Fermi energy is decreased. The peaks near the Fermi energy in the lower energy region are formed due to the nitrogen atoms. The Fermi energy is slightly blueshifted, as in the above case. The conformer $C_{50}AN$-II has the same number

of electrons as the conformer $C_{50}AN$-I. However, in the conformer $C_{50}AN$-II, the nitrogen atom of alanine has broken the one of the C–C bonds of the CNT hexagon and formed two new C–N bonds. This insertion of the nitrogen atom into the C–C bond changes the signatures of the carbon atoms in the band structure of $C_{50}AN$-II (Fig. 8d) at around 0 to −1 eV. Also, the nitrogen and oxygen atoms of the alanine radical introduce new bands below the Fermi energy in the valence band region. In the DOS plot (Fig. 9d), this is reflected in prominent Van Hove singularities.

In the conformer $C_{50}ACOO$-I, a half-filled band appears just above the Fermi energy due to the induced sp^3 defect (Fig. 8e), as in the conformer $C_{50}ANH$-I. Also, a new band appears at −0.6 eV in the valance band region. This is reflected in the DOS plot (Fig. 9e) as a prominent peak just before the Fermi energy, and the DOS at the Fermi energy is slightly increased. In the higher energy region between 1.5 and 2 eV, the DOS is very low and the band gap tends to open up. In the conformer $C_{50}ACO$-I, the removal of OH from the carboxyl group of alanine reduces the number of valence electrons. In the band structure of $C_{50}ACO$-I (Fig. 8f), the Fermi energy has moved up through the conduction band compared to the pristine CNT. A half-filled band originating from the conduction band region passes through the Fermi energy [41]. Thus, the DOS (Fig. 9f) at the Fermi energy is considerably reduced compared to the pristine (5, 0) CNT, and the energy levels are redshifted. The peak adjacent to the Fermi energy in the lower-energy region of the total DOS of pristine CNT is shifted to the higher-energy region in the total DOS of the conformer $C_{50}ACO$-I. The PDOS indicates that this peak is due to the nitrogen and oxygen atoms of the alanine radical. In the lower-energy region at 0.8 eV, there is a peak from the carbon atom of the alanine radical. This can be understood by looking at the geometric structure of conformer $C_{50}ACO$-I, in which the carbon and oxygen atoms of the alanine radical form bonds with the C–C atoms at the bridge site of the CNT.
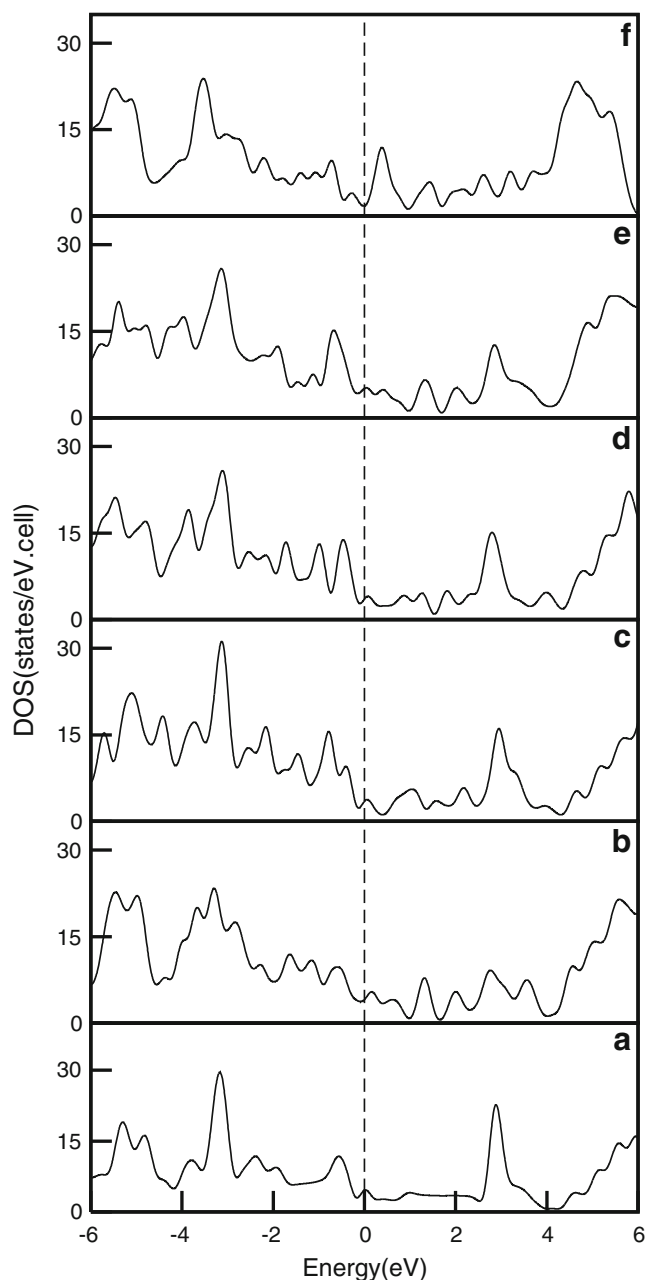
Fig. 9 a–f Density of states plots for a pristine (5, 0) CNT, b $C_{50}ANH$-I, c $C_{50}AN$-I, d $C_{50}AN$-II, e $C_{50}ACOO$-I, f $C_{50}ACO$-I

Summary and conclusions

The noncovalent adsorption of alanine through its functional groups at different orientations onto the surface of a (5, 0) CNT has been systematically investigated. The functional groups (CH_3, NH_2, COOH and OH) of alanine were allowed to face the CNT surface, allowing the CNT π-electron cloud to interact with alanine through XH···π (X=C, N and O) interactions. The binding strength of the alanine with CNT depends on the orientation, the hydrogen bonds and the distance between the alanine and CNT surface. In a study of the interaction of alanine with benzene, OH···π interactions

are reported to be stronger than NH···π and CH···π interactions [42]. However, in our case, we found that interactions of alanine with the CNT through the amine group (NH···π interactions) are more favorable than those through the CH_3, COOH and OH groups. This can be understood by considering the adsorption geometries of the complexes and the electron donor–acceptor interactions between the alanine and the CNT (Table 1). Electrostatic, charge transfer, and van der Waals interactions all play important roles in the stabilization of the conformer. Electronic structure analysis shows an overall blueshift in the energy levels and an increase in the conductivity of the (5, 0) CNT due to the adsorption of alanine.

Covalent functionalization of the CNT was attempted with five different alanine radicals. The interaction of alanine with the CNT through the nitrogen is stronger than the interactions seen for the other conformers. The adsorption of alanine onto CNT (C_{50}ANH-I, C_{50}AN-I, C_{50}ACOO-I, C_{50}ACO-I) modifies the sp^2 hybridization of the CNT structure at the adsorption site, and these particular carbon atoms in the CNT become four coordinated. This modification induces a change in the metallic behavior of the CNT. In the band structure, there is a half-filled band near the Fermi energy due to the sp^3-induced defect level. The adsorption of alanine through the nitrogen atom at the C–C bridge site (C_{50}AN-I) yields the most stable conformer, with two elongated C–N bonds. C_{50}AN-II is the second most stable conformer, and it is the only conformer that retains sp^2 hybridization after functionalization. Similar to what is observed for noncovalent adsorption, energy levels in the DOS plot are blueshifted for the first four conformers. However, for the conformer C_{50}ACO-I, the energy levels are redshifted and the DOS at the Fermi energy is considerably reduced.

Therefore, based on the above investigation, we now have an amino acid functionalized CNT in which the amino acid is physisorbed or chemisorbed with a binding energy that depends on the functional group. This study has opened up a new path for the selection of amino acid functionalized CNTs for biodevices, and these biocompatible nanomaterials are finding extensive application in drug and gene transport. The results of this investigation will be helpful to those attempting to design amino acid functionalized CNTs for various applications.

References

1. Bradley K, Briman M, Star A, Gruner G (2004) Charge transfer from adsorbed proteins. Nano Lett 4:253–256
2. Kuang Z, Kim SN, Crookes-Goodson WJ, Farmer BL, Naik RR (2010) Biomimetic chemosensor: designing peptide recognition elements for surface functionalization of carbon nanotube field effect transistors. ACS Nano 4:452–458
3. Azamian BR, Davis JJ, Coleman KS, Bagshaw CB, Green MLH (2002) Bioelectrochemical single-walled carbon nanotubes. J Am Chem Soc 124:12664–12665
4. Vardanega D, Picaud F, Girarder C (2007) Chiral response of single walled carbon nanotube based sensors to adsorption of amino acids: a theoretical model. J Chem Phys 127:194702–194712
5. Mora MF, Giacomelli CE, Garcia CD (2009) Interaction of L-amino acid oxidase with carbon nanotubes: implications in the design of biosensors. Anal Chem 81:1016–1022
6. Abadir GB, Walus K, Pulfrey DL (2010) Bias-dependent amino-acid-induced conductance changes in short semi-metallic carbon nanotubes. Nanotechnolgy 21:015202–015208
7. Robinson JT, Welsher K, Tabakman SM, Sherlock SP, Wang H, Luong R, Dai H (2010) High performance in vivo near-IR (>1 μm) imaging and photothermal cancer therapy with carbon nanotubes. Nano Res 3:779–793
8. McCarroll J, Baigude H, Yang C, Rana TM (2010) Nanotubes functionalized with lipids and natural amino acid dendrimers: a new strategy to create nanomaterials for delivering systemic RNAi. Bioconjug Chem 21:56–63
9. Lu Q, Moore JM, Huang G, Mount AS, Rao AM (2004) RNA polymer translocation with single-walled carbon nanotubes. Nano Lett 4:2473–2477
10. Lee HJ, Park J, Yoon OJ, Kim HW, Lee DY, Kim DH, Lee WB, Lee NE, Boventre JV, Kim SS (2011) Amine-modified single-walled carbon nanotubes protect neurons from injury in a rat stroke model. Nat Nanotechnol 6:121–125
11. Kam NWS, Lui Z, Dai H (2005) Nanotubes as intracellular transporters for proteins and DNA: an investigation of the uptake mechanism and pathway. Angew Chem Int Ed 44:1–6
12. Zorbas V, Smith AL, Xie H, Ortiz-Acevedo A, Dalton AB, Doeckmann GR, Draper RK, Baughman RH, Musselman IH (2005) Importance of aromatic content of peptide/single-walled carbon nanotube interactions. J Am Chem Soc 127:12323–12382
13. Witus LS, Rocha JDR, Yumono VM, Paramonov SE, Weisman RB, Hartgerink JD (2007) Peptides that non-covalently functionalize single-walled carbon nanotubes to give controlled solubility characteristics. J Mater Chem 17:1909–1915
14. Fan W, Zeng J, Zhang R (2009) Quantum mechanical quantification of weakly interacting complexes of peptides with single-walled carbon nanotubes. J Chem Theor Comput 5:2879–2885
15. Wang Y, Ai H (2009) Theoretical insights into the interaction mechanism between proteins and SWCNTs: adsorptions of tripeptides GXG on SWCNTs. J Phys Chem B 113:9620–9627
16. Roman T, Dino WA, Nakanishi H, Kasai H (2006) Amino acid adsorption on single-walled carbon nanotubes. Eur Phys J D38:117–120
17. Ganji MD (2009) Density functional theory based treatment of amino acids adsorption on single-walled carbon nanotubes. Diamond Relat Mater 18:662–668
18. Sun W, Bu Y, Wang Y (2008) Interaction between glycine/glycine radicals and intrinsic/boron-doped (8, 0) single-walled carbon nanotubes: a density functional theory study. J Phys Chem B 112:15442–15449
19. De Leon A, Jalbout AF, Basiuk VA (2008) SWNT–amino acid interactions: a theoretical study. Chem Phys Lett 457:185–190

20. Chang CM, Jalbout AF (2010) Metal induced amino acid adsorption on nanotubes. Thin Solid Films 518:2070–2076

21. Rajesh C, Majumder C, Mizuseki H, Kawazoe Y (2009) A theoretical study on the interaction of aromatic amino acids with graphene and single walled carbon nanotube. J Chem Phys 130:124911–124916

22. Bianco A, Kostaredes K, Partidos CD, Prato M (2005) Biomedical applications of functionalized carbon nanotubes. Chem Commun 571–577

23. Zhang Y, Li J, Shen Y, Wang M, Li J (2004) Poly-L-lysine functionalization of single-walled carbon nanotube. J Phys Chem B 108:15343–15346

24. Pulikkathara MX, Khabasheska VN (2008) Covalent sidewall functionalization of single-walled carbon nanotubes by amino-acids. Russ Chem Bull Int Edn 57:1054–1062

25. Carneiro MA, Venezuela P (2008) First principles calculations of alanine radicals adsorbed on pristine and functionalized carbon nanotubes. J Phys Chem C 112:14812–14815

26. Kresse G, Furthmuller J (1996a) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169–11186

27. Kresse G, Furthmuller J (1996b) Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Mater Sci 6:15–50

28. Blochl PE (1994a) Projector augmented-wave method. Phys Rev B 50:17953–17979

29. Kresse G, Joubert D (1999) From ultrasoft pseudopotentials to the projector augmented-wave method. Phys Rev B 59:1758–1775

30. Perdew JP, Zunger A (1981) Self-interaction correction to density-functional approximations for many-electron systems. Phys Rev B 23:5048–5079

31. Perdew JP, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Singh DJ, Fiolhais C (1992) Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation. Phys Rev B 46:6671–6687

32. Ortmann F, Schmidt WG, Bechstedt F (2005) Attracted by long range electron correlation: adenine and graphite. Phys Rev Lett 95:186101–186104

33. Lim S, Park N (2009) Ab initio study of noncovalent sidewall functionalization of carbon nanotubes. Appl Phys Lett 95:243110–243113

34. Tournus F, Latil S, Heggie MI, Charlier JC (2005) π-Stacking interaction between carbon nanotubes and organic molecules. Phys Rev B 72:075431–075435

35. Simeonia M, De Luca C, Picozzi S, Santucci S, Delley B (2005) Interaction between zigzag single-wall carbon nanotubes and polymers: a density-functional study. J Chem Phys 122:214710–214718

36. Monkhorst HJ, Pack JD (1976) Special points for Brillouin-zone integrations. Phys Rev B 13:5188–5192

37. Henkelman G, Arnaldsson A, Jonsson H (2006) A fast and robust algorithm for Bader decomposition of charge density. Comput Mater Sci 36:254–360

38. Kokalj A (2005) Computer graphics and graphical user interfaces as tools in simulations of matter at the atomic scale. Comput Mater Sci 28:155–168

39. Gulseren O, Yildirim T, Ciraci S (2002) Systematic ab initio study of curvature effects in carbon nanotubes. Phys Rev B 65:153405–4

40. Iyakutti K, Bodapati A, Peng X, Keblinsi P, Nayak SK (2006) Electronic band structure, electron–phonon interaction, and superconductivity of (5, 5), (10, 10), and (5, 0) carbon nanotubes. Phys Rev B 73:35413–35415

41. Zhao J, Park H, Han J, Lu JP (2004) Electronic properties of carbon nanotubes with covalent sidewall functionalization. J Phys Chem B 108:4227–4230

42. Mohan N, Vijayalakshmi KP, Koga N, Suresh CH (2010) Comparison of aromatic NH–π, OH–π and CH–π interactions of alanine using MP2, CCSD and DFT methods. J Comput Chem 16:2874–2882

ORIGINAL PAPER

Density functional theory study of small nickel clusters

Satyender Goel · Artem E. Masunov

Abstract The stable geometries and atomization energies for the clusters Ni_n ($n=2–5$) are predicted with all-electron density functional theory (DFT), using the BMK hybrid functional and a Gaussian basis set. Possible isomers and several spin states of these nickel clusters are considered systematically. The ground spin state and the lowest energy isomers are identified for each cluster size. The results are compared to available experimental and other theoretical data. The molecular orbitals of the largest cluster are plotted for all spin states. The relative stabilities of these states are interpreted in terms of superatom orbitals and no-pair bonding.

Keywords Density functional theory · Hybrid exchange-correlation functionals · Unrestricted Kohn–Sham · Molecular orbitals · Geometry and energetics · Small metallic clusters

Introduction

The interest in transition metal clusters is based on their numerous applications, many of which are related to

S. Goel (✉)
Department of Chemical and Biological Engineering,
Northwestern University,
2145 Sheridan Road,
Evanston, IL 60208, USA
e-mail: satyendergoel@gmail.com

A. E. Masunov (✉)
NanoScience Technology Center and Department of Chemistry,
University of Central Florida,
12424 Research Parkway, Suite 400,
Orlando, FL 32826, USA
e-mail: amasunov@mail.ucf.edu

catalysis. Recently, nanocatalysts begin to emerge as a promising alternative, combining the advantages of both homo- and heterogeneous catalysts and often exhibiting unique activity. The properties of nanocatalysis are strongly dependent on their dimensionality, size, chemical composition and the morphology of the active nanoparticles. Considerable efforts have been directed in recent decades [1–3] at the design and controlled fabrication of nanostructured catalytic materials. Metal clusters represent an important class of nanocatalysts with several unique properties. Their ability to act as catalysts is determined by several factors: (i) a high surface to volume ratio; (ii) the availability of active absorption and reaction sites; (iii) low energy barriers to restructuring; (iv) a wide range of coordination numbers; and (v) easy migration and interatomic rearrangements that facilitate bond breaking.

Transition metal nanoclusters represent a particular type of system intermediate between molecules and solids. Unlike molecules and solids, the theory of metal nanoclusters lacks a set of simple rules that can assist in predicting stable structures, although such rules have begun to emerge recently [4, 5]. Thus, simulations based on first principles remain the only method that can be used to predict their geometries at present. Studies of transition metal (TM) systems present a challenge for their theoretical description, due to the presence of several closely spaced energy levels, which result in strong electron correlation [6–8]. For this reason, molecules and clusters containing TMs require careful selection of the theoretical methods used. This contribution aims to provide an accurate theoretical description of small clusters of nickel ($Ni_{2–5}$) using density functional theory methods.

Among $3d$ TM clusters, nickel systems have received the most attention, both experimentally and theoretically [6–

20]. In the past three decades, small nickel clusters and their properties have attracted great interest from researchers. The chemical, physical, electronic and magnetic properties of nickel nanoclusters are related to their geometries, and play an important role in understanding the transition from diatomics to the bulk. Although the geometries of small nickel clusters hold the key to their behavior, studying their shapes experimentally presents a challenge. Luckily, several experimental studies have published on Ni_{2-5} clusters [10, 14, 17, 18, 21], and the results from these can be used for theoretical comparison.

A significant number of theoretical studies have been performed on small nickel clusters [3, 6–9, 12, 13, 15, 16, 21–23]. Recently, Arvizu et al. [9] reported several different topologies of small nickel clusters (Ni_{2-5}) in their pure DFT study. They determined structural properties, binding energies, and adiabatic ionization potentials. To address the issue of the ground-state multiplicities of these nickel clusters, they tested five spin states (singlet, triplet, quintet, septet, nonet) for each of the clusters. Other DFT predictions for the structural parameters of these clusters, as obtained using LDA [19, 20] and GGA [13] methods, have also been reported. Petkov et al. [23] presented the effect of an impurity on the structure, stability, electronic and magnetic properties of Ni_4, and discussed six different symmetries of the bare Ni_4 cluster: two planar (D_{2h} and C_{2v}) and four nonplanar (C_{3v}, T_d, C_{2v}, C_s). They reported that C_s is the most stable geometry for Ni_4, and found Jahn–Teller distortion in Ni_4 with T_d symmetry. In another study of the structural and magnetic properties of nickel clusters (Ni_{2-13}), Xie et al. [24] discussed instabilities due to Jahn–Teller distortion in high-symmetry clusters possessing multiple degeneracy. They also reported new ground-state structures for Ni_5 and Ni_7 with C_{2v} and C_2 symmetry, respectively.

Aside from DFT, more approximate techniques have also been employed to investigate small nickel clusters, including density functional tight binding (DFTB) [12], the embedded-atom method (EAM) [22], and effective medium theory (EMT) [15]. Considering the existence of multiple minima on the potential energy surfaces of transition metal clusters, DFTB may be a better alternative to DFT, as it provides a compromise between accuracy and computational speed, and thus the opportunity for extensive geometrical searches. DFTB was used in a molecular dynamics (MD) simulated annealing technique employed by Luo [12]. On the other hand, Nygren et al. [15] reported binding energies for Ni_{4-9} obtained in an EMT study. They described the ionization potentials of these Ni clusters using two different approaches (at the one-electron ECP level, and an all-electron description) to explain the odd-even effect observed experimentally. Grigoryan et al. [22] reported the four lowest-energy isomers among all of the

clusters from 2 to 150, and determined structural information for each individual cluster. This study was done with an EAM method that was specially developed for studying metal interactions.

In this contribution, we investigate the geometry and energetics of small nickel clusters (Ni_{2-5}), as shown in Fig. 1, with a modern BMK (Boese–Martin for kinetics) hybrid meta-GGA functional [25]. We consider all multiplicities possible for these clusters, ranging from singlet to tridectet. Here, we present our prediction of the lowest energy state and compare our results with existing experimental and theoretical data. We have also studied various isomers of Ni_4 to determine its ground-state geometry.

Computational details

All calculations were done with the Gaussian 2003 [26] program using an all-electron Wachters+f [27, 28] basis set. The initial guess was generated using a Harris functional [29], which is the default option in Gaussian 2003. Spin-polarized (unrestricted) ansatz DFT and the BMK exchange-correlation functional were used throughout. BMK is a hybrid meta-GGA functional that is designed to accurately describe transition-state (TS) properties as well as the atomization energies, geometries and harmonic frequencies of molecules in the ground state. A BMK functional with a large fraction of Hartree–Fock exchange (42%) was developed based on a diverse and balanced parameterization set that included transition metal complexes and hydrogen-bonded systems [25]. Previously, we found BMK to be superior for small systems containing Ni atoms, such as TM hydrides [30] and TM carbide diatomics [31].

To avoid false SCF solutions with Fermi holes (i.e., virtual orbitals with energies lower than some of the occupied ones), a special SCF algorithm was employed that made use of the fractional occupation numbers (FON) around the Fermi energy. This was accomplished by using the keywords SCF=Fermi and IOp(5/22=5). In the FON approach, the orbital occupations are determined using the Fermi–Dirac function for a fictitious electron temperature, so that the sum of the occupation numbers equals the correct number of electrons for the system [32]. The electron temperature value is set to 3000 K for the initial SCF cycle, and then lowered to 0 K in ten SCF steps, so that the occupational numbers become integers during the final SCF cycles. The stability of the SCF solution was checked and the KS orbitals were reoptimized (if unstable) using the keyword Stable=Opt. The default integration grid was used for all calculations. The GaussView graphical interface was used to examine the Kohn–Sham orbitals.

Fig. 1 Geometries of nickel clusters (Ni$_{2-5}$). The four isomers of Ni$_4$ are labeled with indices in parentheses: Ni$_4$(s) is square planar, Ni$_4$(t1) is tetrahedral, Ni$_4$(t2) is distorted tetrahedral, and Ni$_4$(r) is rhombic
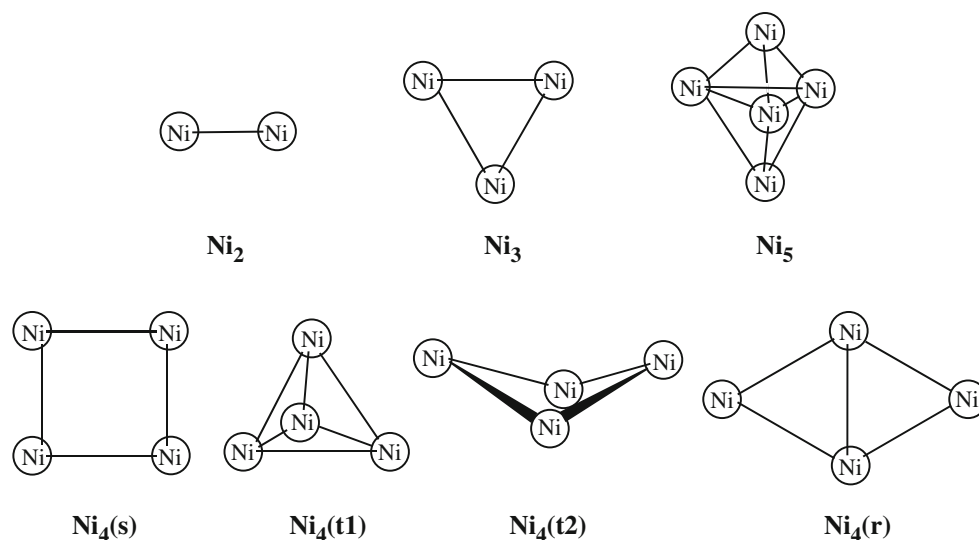
Results and discussion

The optimized geometries of small nickel clusters (Ni$_{2-5}$) are reported in Table 1. This includes all possible spin multiplicities of small nickel clusters, starting from the lowest at the top to the highest at the bottom, and for Ni$_2$ on the left through to Ni$_5$ on the right. We considered the four lowest energy isomers of Ni$_4$ in order to analyze the structural changes and variation in binding energy. To the best of our knowledge, this study provides the first comprehensive analysis of all possible spin states of small nickel clusters (Ni$_{2-5}$). As discussed in detail in the following sections, our results indicate that the BMK-predicted energies and geometries are in good agreement with experimental results (where available), and provide a good benchmark for further investigations of systems containing small Ni clusters. All predicted energetic and geometric properties for the various spin states of small Ni clusters are presented in Table 1. We compare our results with the limited theoretical and experimental data in the following sections.

Diatomic Ni

Table 1 reports the bond length variations in Ni$_2$ from singlet through to septet; the latter was found to be unstable with respect to dissociation into triplet atoms. We predict that the ground electronic state for Ni$_2$ is a Boltzmann mixture of S and T with binding energies of 1.99 eV and 1.98 eV, respectively. The singlet was more stable by 0.01 eV than the triplet, in agreement with the results of earlier experimental studies [10, 14, 17, 18, 21, 33]. Our bond lengths for S and T (2.303 Å and 2.316 Å, respectively) are both longer then the experimental value of 2.155 Å [18], while the binding energies are in good agreement with
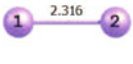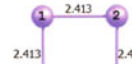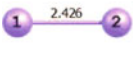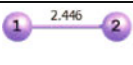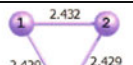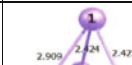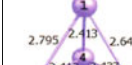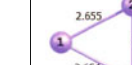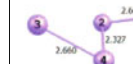
the experimental value of 2.1 eV. Several DFT studies have been performed in the past using pure DFT (LDA and GGA) functionals [9, 13, 16, 19, 20, 23, 24], which are reported to produce better geometries but not energies. Similarly, our BMK predictions for triplet and singlet dissociation energies are better than ab initio CASPT2 [21] ones (1.93 eV and 1.95 eV, respectively), while the CASPT2 predictions for the bond lengths (2.24 Å and 2.25 Å) [21] are superior to our BMK predictions. It is worth noting the increase in bond length and decrease in binding energy with increasing multiplicity, which disfavors higher spin states due to easy bond breaking.

Triatomic Ni

The structures of Ni$_3$ in the singlet, triplet, quintet, septet and nonet states are reported in Table 1. We obtained different geometries for various multiplicities, where the triplet and quintet had triangular geometries with three different sides, while the septet and nonet were isosceles. The singlet converges to a bent geometry with one angle >120°. We found that two states, the triplet and septet, were similar in energy (3.01 eV and 3.09 eV) and more stable than the singlet, quintet and nonet. Structural studies of Ni$_3$ are often considered to be of interest due to a disagreement between theory and experiment. On the one hand, C$_{2v}$ symmetry was found experimentally [24], with an apex angle of >60°; on the other hand, tight binding potential based MD [12] and EAM [22] simulations have predicted Ni$_3$ to have D$_{3h}$ (equilateral triangle) symmetry. LDA studies [24] gave an isosceles triangle type geometry. Arvizu et al. [9] reported that the triplet is the ground state for Ni$_3$, with an isosceles geometry, in their DFT study of five spin states for Ni$_3$. Although they used different initial structures for Ni$_3$, their optimized geometry was still isosceles, with all angles less

Table 1 Geometries of Ni_2, Ni_3, Ni_5 and four isomers of Ni_4 with several spin states (singlet, triplet, quintet, septet, nonet, undectet, duodectet and tridectet), along with their bond lengths (Å) and binding energies (eV), as displayed by each structure. Binding energies in parentheses are the last SCF energies after several steps for unconverged geometries

M	Ni_2	Ni_3	$Ni_4(s)$	$Ni_4(t1)$	$Ni_4(t2)$	$Ni_4(r)$	Ni_5
1	1.99	2.92	0.21	4.48	4.50	5.42	7.17
3	1.98	3.01	4.37	4.48	4.49	5.43	6.37
5	0.95	2.56	5.43	4.57	5.39	4.76	7.17
7	-1.72	3.09	4.53	4.59	4.61	4.84	7.25
9		2.31	2.84	3.80	3.80	3.05	6.74
11				(1.00)	0.95	1.02	(4.96)
13				-1.89	-1.77	-1.39	

than 90°, which is in agreement with our findings. An LDA and GGA study with various basis sets [9] reported only the energies of the lowest spin state (triplet). The best binding energy reported by Arvizu et. al. [9] with TZVP-GGA (3.32 eV) is more than our BMK predictions. The higher energy in this case is likely to be due to overestimation on the part of the GGA method, consistent with the Ni dimer results (2.64 eV from [9] vs. our BMK predictions of 1.99 eV, as compared to the experimental value of 2.1 eV [18]).

Tetratomic Ni
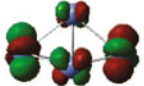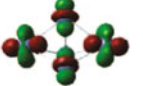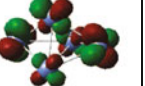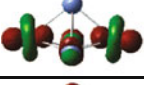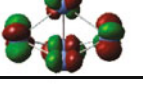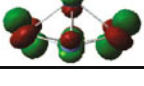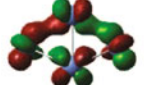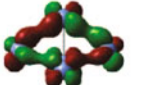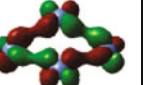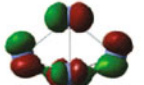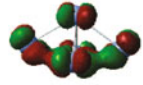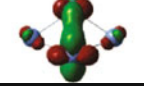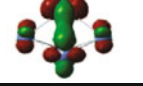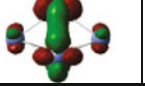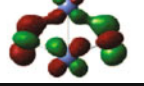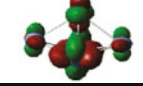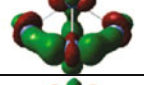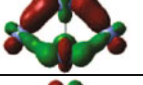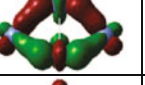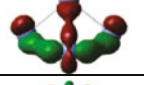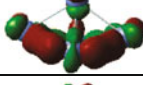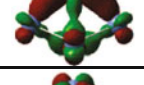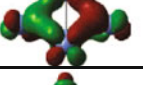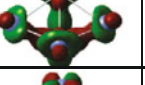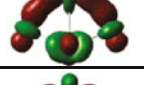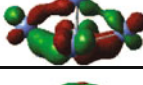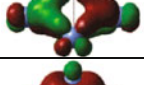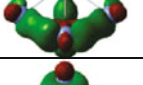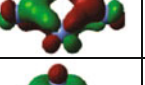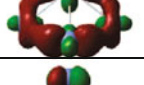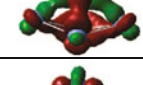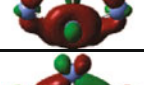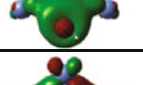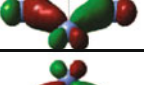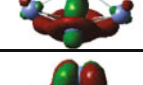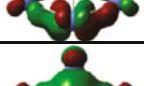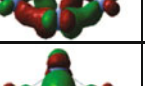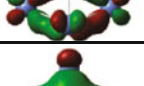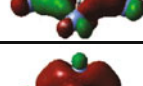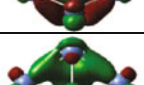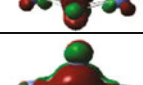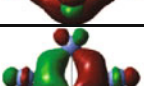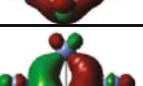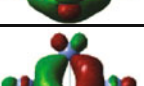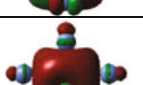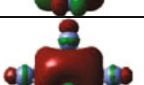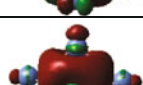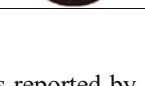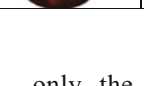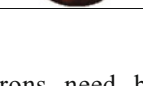
The results for various low-lying isomers of Ni_4 in several spin states are reported in Table 1. We were unable to find any experimental data for Ni_4 to make comparison with our findings. Full geometry optimization followed by vibrational analysis was performed on different initial geometries

(square planar, two tetrahedral, and rhombic) for Ni_4 in several spin states (singlet, triplet, quintet, septet, nonet, undectet and tridectet). We represent these structures with a letter in parentheses: $Ni_4(s)$ is square planar, $Ni_4(t1)$ is tetrahedral, $Ni_4(t2)$ is distorted tetrahedral, and $Ni_4(r)$ is rhombic. We determined that the rhombic (planar) structure was favored in the case of Ni_4 in all but one nonet state, where $Ni_4(t1)$ and $Ni_4(t2)$ were equally stable. We also found that many of our other initial geometries [square planar $^5Ni_4(s)$ and distorted tetrahedral $^5Ni_4(t2)$] converged to the stable rhombic (planar) type structure $Ni_4(r)$. This led us to draw the conclusion that the quintet is the most stable spin state for Ni_4. Our results differ from those reported by Arvizu et al. [9] and Petkov et al. [23]. They reported that quintet $Ni_4(t1)$ is the ground state structure, while we predict that quintet $Ni_4(r)$ is the ground-state geometry. The binding energy of $^5Ni_4(t1)$, 4.02 eV (GGA/TZVP level of

Table 2 Molecular orbitals of all stable spin states (singlet, triplet, quintet, septet and nonet) of Ni$_5$, tabulated as per their distributions, with antibonding (p orbitals) on top, then unoccupied, nonbonding orbitals (lone pairs, d orbitals), and bonding s orbitals at the bottom

Ni$_5$_09		Ni$_5$_07		Ni$_5$_05		Ni$_5$_03		Ni$_5$_01	
6.74		**7.25**		**7.17**		**6.37**		**7.17**	
L+4/79 -0.060		-	-	-	-	-	-	-	-
L+3/78 -0.077		L+4/78 -0.075		-	-	-	-	-	-
L+2/77 -0.085		L+3/77 -0.088		L+4/77 -0.085		-	-	-	-
L+1/76 -0.087		L+2/76 -0.088		L+3/76 -0.085		L+4/76 -0.087		-	-
L/75 -0.116		L+1/75 -0.117		L+2/75 -0.119		L+3/75 -0.112		L+4/75 -0.118	
S/74 -0.176		L/74 -0.170		L+1/74 -0.172		L+2/74 -0.181		L+3/74 -0.171	
S-1/73 -0.183		S/73 -0.187		L/73 -0.183		L+1/73 -0.183		L+2/73 -0.185	
S-2/72 -0.215		S-1/72 -0.215		S/72 -0.218		L/72 -0.210		L+1/72 -0.216	
S-3/71 -0.250		S-2/71 -0.248		S-1/71 -0.249		S/71 -0.252		L/71 -0.248	
S-4/70 -0.252		S-3/70 -0.250		S-2/70 -0.250		S-1/70 -0.252		H/70 -0.250	
S-5/69 -0.255		S-4/69 -0.253		S-3/69 -0.255		H/69 -0.252		H-1/69 -0.253	
S-6/68 -0.255		S-5/68 -0.255		H/68 -0.255		H-1/68 -0.253		H-2/68 -0.255	
S-7/67 -0.255		H/67 -0.258		H-1/67 -0.257		H-2/67 -0.259		H-3/67 -0.257	
H/66 -0.258		H-1/66 -0.259		H-2/66 -0.260		H-3/66 -0.261		-	-

Table 2 (continued)

H-1/65 −0.261		H-2/65 −0.261		H-3/65 −0.262		-	-	-	-
H-2/64 −0.264		H-3/64 −0.262		-	-	-	-	-	-
H-3/63 −0.266		-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-
56 −0.289	56 −0.285	56 −0.284	56 −0.288	56 −0.285	
55 −0.294	55 −0.289	55 −0.293	55 −0.289	55 −0.290	
54 −0.298	54 −0.300	54 −0.302	54 −0.301	54 −0.300	
53 −0.308	53 −0.300	53 −0.304	53 −0.302	53 −0.301	
52 −0.310	52 −0.304	52 −0.304	52 −0.309	52 −0.304	
51 −0.314	51 −0.310	51 −0.314	51 −0.313	51 −0.310	
50 −0.314	50 −0.316	50 −0.316	50 −0.314	50 −0.315	
49 −0.324	49 −0.319	49 −0.321	49 −0.325	49 −0.319	
48 −0.330	48 −0.334	48 −0.333	48 −0.326	48 −0.333	
47 −0.357	47 −0.355	47 −0.360	47 −0.347	47 −0.356	
46 −0.436	46 −0.431	46 −0.436	46 −0.434	46 −0.432	

theory), as reported by Arvizu et. al. [9], was less than that predicted in this study (4.57 eV) for the same spin state and geometry, which differed from the observations made in the case of Ni$_3$ above. In another study by Reuse et al. [20], the authors pointed out the importance of Jahn–Teller distortion in predicting the correct structure. They stated that for the atomic Ni configuration d^9s^1, d electrons can be ignored; only the contribution from the s electrons need be considered. According to the Jahn–Teller theorem, an electronic system with degenerate energy levels is unstable, and structural distortion will occur in order to remove this degeneracy [34]. This is why we observed that the square-planar tetramer with four valent electrons underwent structural distortion into the planar rhombus structure.

Nygren et al. [15] found that planar D_{2h} was the lowest energy isomer of all five isomers they tested. Table 1 shows that the higher spin states are unstable for Ni_4 isomers, and undectet is loosely bound with a binding energy that is four times smaller than those of the lower spin states.

Pentatomic Ni

We investigated one lowest energy structural isomer of Ni_5, which was reported in earlier studies (both experimental and theoretical) [9, 13, 17, 19, 20] to have trigonal bipyramidal geometry. Several spin states (reported in Table 1) were optimized to a symmetric conformation (D_{3h}) in all cases, except in the case of undectet, where no convergence was achieved. We determined that $^7Ni_5$ is the ground-state structure, whereas $^1Ni_5$ and $^5Ni_5$ are very close in energy, which is common in transition metal clusters. The primary factor contributing to this multitude character of states was the participation of both s and d orbitals in bonding. Michelini et al. [13] reported the existence of two stable quasi-degenerate isomers, the septet (C_{4v} symmetry) and quintet (D_{3h} symmetry), for Ni_5. Their study was supported by the recent [9] investigations of five different topologies of the Ni_5 cluster obtained with VWN (LDA) calculations, which all converged to three-dimensional structures, with the quintet bipyramidal state being the lowest, followed by the septet pyramidal structure. They noted state inversion with GGA functionals, where the pyramidal structure in the septet state was the ground state. None of our Ni_5 spin states converged to a pyramidal geometry.

Table 2 displays the molecular orbitals of all spin states (singlet, triplet, quintet, septet and nonet) for the Ni_5 pentamer. This picture can be rationalized using the cluster (or superatom) orbital concept [35]. Table 2 shows that the lowest molecular orbital (46 in all multiplicities) is a cluster s orbital, a fully symmetric combination of atomic p orbitals. Next, there are 25 combinations of atomic d orbitals into nonbonding or lone pair orbitals. These are nearly degenerate and singly occupied in the most stable spin multiplicity (septet), in agreement with Hund's rule. Three cluster p orbitals are found at higher energy. These are antibonding with respect to some of the cluster atoms. Table 2 also lists the geometry of the respective spin state at the top of the table along with the binding energy. The higher multiplicities are stable (although not as stable as the septet) due to "no-pair bonding" [36], the electronic effect that holds together the one-electron system H_2^+.

Conclusions

We investigated the geometries and energetics of small ($n = 2$–5) nickel clusters using a broken-symmetry DFT theory approach. The hybrid BMK exchange-correlation functional with an all-electron basis set performed well in this study. Selected isomers in all possible spin states were investigated for these four clusters. An exhaustive search for the geometric isomers, which is presently feasible for tetratomic systems [37], remains the subject of future investigation. All of the minima found were characterized by a stability analysis of the Kohn–Sham determinant in order to ensure the accuracy of the ground spin state of the respective cluster. The trends obtained for the energetic and structural properties experimentally were well reproduced in the present work. We found that the BMK functional was suitable for predicting transition metal cluster properties, and this method can be used to study larger nickel clusters as well as more complicated transition metal systems.

References

1. Ovsitser O, Kondratenko EV (2009) Similarity and differences in the oxidative dehydrogenation of C_2–C_4 alkanes over nano-sized VO_x species using N_2O and O_2. Catal Today 142:138–142. doi:10.1016/j.cattod.2008.09.012
2. Uddin J, Morales CM, Maynard JH, Landis CR (2006) Computational studies of metal–ligand bond enthalpies across the transition metal series. Organometallics 25:5566–5581. doi:10.1021/om0603058
3. Simoes JAM, Beauchamp JL (1990) Transition metal hydrogen and metal carbon bond strengths: the keys to catalysis. Chem Rev 90:629–688
4. Zubarev DY, Boldyrev AI (2008) Developing paradigms of chemical bonding: adaptive natural density partitioning. Phys Chem Chem Phys 10:5207–5217. doi:10.1039/b804083d
5. Zubarev DY, Boldyrev AI (2009) Deciphering chemical bonding in golden cages. J Phys Chem A 113:866–868. doi:10.1021/jp808103t
6. Jensen KP, Roos BO, Ryde U (2007) Performance of density functionals for first row transition metal systems. J Chem Phys 126:014103–014116
7. Furche F, Perdew JP (2006) The performance of semilocal and hybrid density functionals in $3d$ transition-metal chemistry. J Chem Phys 124:044103–044129
8. Harrison JG (1983) Density functional calculations for atoms in the 1st transition Series. J Chem Phys 79:2265–2269
9. Arvizu GL, Calaminici P (2007) Assessment of density functional theory optimized basis sets for gradient corrected functionals to transition metal systems: the case of small Ni_n ($n \leq 5$) clusters. J Chem Phys 126:194102–194111. doi:19410210.1063/1.2735311
10. Kant A (1964) Dissociation energies of diatomic molecules of the transition elements. I. Nickel. J Chem Phys 41:1872–1876
11. Knickelbein MB, Yang S, Riley SJ (1990) Near-threshold photoionization of nickel clusters: ionization potentials for Ni_3 to Ni_{90}. J Chem Phys 93:94–104

12. Luo CL (2000) The structure of small nickel clusters: Ni_2–Ni_{19}. Model Simul Mater Sci Eng 8:95–101

13. Michelini MC, Diez RP, Jubert AH (2004) Density functional study of the ionization potentials and electron affinities of small Ni_n clusters with n=2–6 and 8. Comput Mater Sci 31:292–298. doi:10.1016/j.commatsci.2004.03.018

14. Moskovits M, Hulse JE (1977) Ultraviolet-visible spectra of diatomic, triatomic, and higher nickel clusters. J Chem Phys 66:3988–3994

15. Nygren MA, Siegbahn PEM, Wahlgren U, Akeby H (1992) Theoretical ionization energies and geometries for Ni_N ($4 \leq N \leq 9$). J Phys Chem 96:3633–3640

16. Onal I, Sayar A, Uzun A, Ozkar S (2009) A density functional study of Ni_2 and Ni_{13} nanoclusters. J Comput Theor Nanos 6:867–872. doi:10.1166/jctn.2009.1119

17. Parks EK, Zhu L, Ho J, Riley SJ (1994) The structure of small nickel clusters Ni_3–Ni_{15}. J Chem Phys 100:7206–7222

18. Pinegar JC, Langenberg JD, Arrington CA, Spain EM, Morse MD (1995) Ni_2 revisited: reassignment of the ground electronic state. J Chem Phys 102:666–674

19. Reuse FA, Khanna SN (1995) Geometry, electronic-structure, and magnetism of small Ni_N (N=2–6, 8, 13) clusters. Chem Phys Lett 234:77–81

20. Reuse FA, Khanna SN (1999) Photoabsorption spectrum of small Ni_n (n=2–6, 13) clusters. Eur Phys J D 6:77–81

21. Pouamerigo R, Merchan M, Nebotgil I, Malmqvist PA, Roos BO (1994) The chemical-bonds in CuH, Cu2, NiH, and Ni_2 studied with multiconfigurational 2nd-order perturbation theory. J Chem Phys 101:4893–4902

22. Grigoryan VG, Springborg M (2004) Structural and energetic properties of nickel clusters: $2 \leq N \leq 150$. Phys Rev B 70:205415–205429. doi:10.1103/PhysRevB.70.205415

23. St Petkov P, Vayssilov GN, Kruger S, Rosch N (2006) Structure, stability, electronic and magnetic properties of Ni_4 clusters containing impurity atoms. Phys Chem Chem Phys 8:1282–1291. doi:10.1039/b518175e

24. Xie Z, Ma QM, Liu Y, Li YC (2005) First-principles study of the stability and Jahn–Teller distortion of nickel clusters. Phys Lett A 342:459–467. doi:10.1016/j.physleta.2005.05.067

25. Boese AD, Martin JML (2004) Development of density functionals for thermochemical kinetics. J Chem Phys 121:3405–3416

26. Frisch MJ et al. (1994–2003, 2004) Gaussian 03, revision D.01. Gaussian Inc., Wallingford

27. Wachters AJ (1970) Gaussian basis set for molecular wavefunctions containing third-row atoms. J Chem Phys 52:1033–1038

28. Hay PJ (1977) Gaussian basis sets for molecular calculations—representation of $3d$ orbitals in transition-metal atoms. J Chem Phys 66:4377–4384

29. Harris J (1985) Simplified method for calculating the energy of weakly interacting fragments. Phys Rev B 31:1770–1779

30. Goel S, Masunov AE (2008) Potential energy curves and electronic structure of $3d$ transition metal hydrides and their cations. J Chem Phys 129:214302–14

31. Goel S, Masunov AE (2008) First-principles study of transition metal diatomics as the first step in multiscale simulations of carbon nanotube growth process. In: 4th Int Conf on Multiscale Material Modeling, Tallahassee, FL, USA, 27–31 Oct 2008, pp 110–113

32. Rabuck AD, Scuseria GE (1999) Improving self-consistent field convergence by varying occupation numbers. J Chem Phys 110:695–700

33. Wang HM, Haouari H, Craig R, Lombardi JR, Lindsay DM (1996) Raman spectra of mass-selected nickel dimers in argon matrices. J Chem Phys 104:3420–3422

34. Ham FS (2000) The Jahn–Teller effect: a retrospective view. J Lumin 85:193–197

35. Jiang DE, Whetten RL, Luo WD, Dai S (2009) The smallest thiolated gold superatom complexes. J Phys Chem C 113:17291–17295. doi:10.1021/jp9035937

36. Monari A, Pitarch-Ruiz J, Bendazzoli GL, Evangelisti S, Sanchez-Marin J (2010) High-spin states in tetrahedral X_4 clusters (X=H, Li, Na, K). Int J Quantum Chem 110:874–884. doi:10.1002/qua.21987

37. Olson JK, Boldyrev AI (2009) Ab initio search for global minimum structures of the novel B_3H_y (y=4–7) neutral and anionic clusters. Inorg Chem 48:10060–10067. doi:10.1021/ic900905h

Effect of the methylation of uracil and/or glycine on their mutual interaction

Hongqi Ai · Dejie Li · Yongping Zhao · Chong Zhang · Qiang Li · Jijun Feng

Abstract In order to simulate the hydrogen bonding and proton transfer (PT) in protein-DNA/RNA interactions, a series of simplified models were employed and investigated in the gas phase. These models included various neutral, anionic and cationic glycine-uracil dimers, and their methylated derivatives generated by the mono- or dimethylation of glycine and/or uracil moieties of the dimer. The results reveal that the only process that can occur in the neutral complexes is a double-PT process leading to proton exchange between the two moieties (i.e., point mutation). The first methyl substitute can reduce the activation energy of the PT process and thus promote the isomerization of the two moieties; further methylation can reduce the isomerization in only some of the cases. In the anionic complexes, only the one-way PT (i.e., amino acid → nucleic acid base) process is energetically favorable, and this PT process is an interesting barrier-free one (BFPT), with the attached electron locating itself at the base moiety. Methylation will disfavor BFPT, but it cannot alter the nature of BFPT. In the cationic complexes, three different PT processes can occur. These processes can transform mutually by adjusting either or both of the methylated sites and methyl number, indicating that the methylation can regulate the dynamics of these PT processes.

Keywords Methyl effect · Proton transfer · Charged dimers

H. Ai (✉) · D. Li · Y. Zhao · J. Feng
Shandong Provincial Key Laboratory of Fluorine Chemistry and Chemical Materials, School of Chemistry and Chemical Engineering, University of Jinan,
Jinan 250022, 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, Shandong Province, People's Republic of China

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

Introduction

Rare imino/enol tautomeric transformations of a nucleic acid base (often simply termed a "base") can occur via the concerted transfer of two protons in Watson–Crick bases, inducing point mutations during the replication of DNA or RNA [1, 2]. For the adenine–thymine pair, such a process would lead to a thymine tautomerization, and thus the mispairing of the bases [3, 4]. For the uracil tautomerization process, this can also be achieved by intra- or intermolecular proton transfer (PT). Ultimately, the "normal" uracil (see Fig. 1) will transform into a rarer one [5]. These rare tautomers of bases are generally less stable than the normal form, but can be stabilized by interacting with cellular components such as amino acids or ions. For example, phenylalanine is conserved among all the uracil DNA glycosidases for which sequences are known [6]. Among all of these DNA glycosidases, it is a common structural feature for an aromatic enzyme side chain to stack with a damaged base. Obviously, this side chain stacking stabilizes the ionized base. Alanine mutation, however, would sacrifice the stacking interactions with uracil and reduce the stability of the nucleotide-flipped complex [7]. One structural difference between alanine and phenylalanine is their side chains. Thus, we can regulate the stability of the complex by substituting
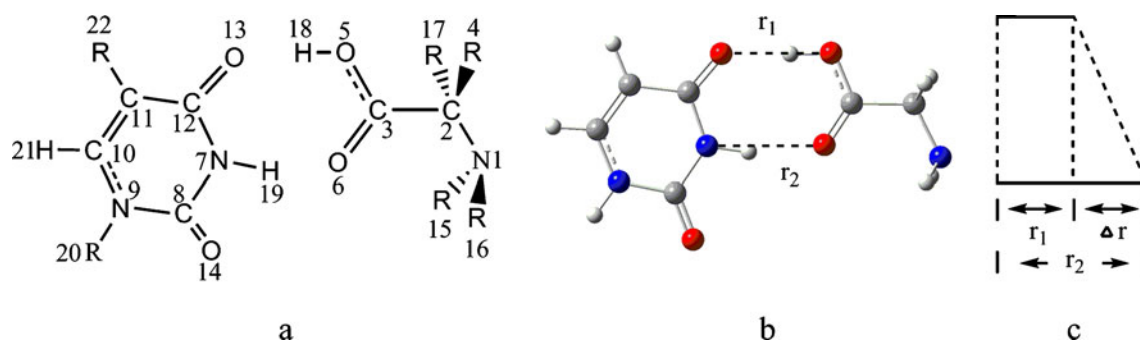
Fig. 1a–c Interaction mode (**a**) of U and G (R = H, CH₃), and the hydrogen bonds involved (**a, b, c**)

different amino acid residues at the phenylalanine sites of the uracil DNA glycosidase sequences. Studies from Luscome et al. [8] revealed that only glycine among the nonpolar amino acids coupled with DNA, and that alanine and glycine with smaller side chains in the hydrophobic residues could also form bonds with DNA using their main-chain atoms. The hydrogen bonding between the carboxyl oxygen of an amino acid residue and the bases of DNA/RNA provides the greatest specificity [9], and thus represents an important interaction among protein-DNA/RNA interactions.

To investigate the stability of a mutagenic tautomer of uracil induced by amino acids, Dabkowska et al. [3, 4] studied the interaction between uracil and glycine at the density functional B3LYP/6-31++G** level. Their results revealed that the most stable uracil–glycine dimer (UG1) was a cyclic structure with the H18 and O6 atoms of the most stable canonical glycine hydrogen bonding with the O14 and H20 atoms of the most stable uracil, respectively. In vivo, however, this interaction mode is not available as sugar substitutes for the H20 of the uracil. Therefore, the UG shown in Fig. 1a—the second most stable dimer [4]—is the most available one. Moreover, the anionic UG form is also the most stable one among its anionic isomers. Once the anionic UG is formed or an electron attaches to the neutral UG dimer, the proton of the glycine moiety will transfer to the uracil moiety spontaneously. For the most stable UG1, PT does not take place spontaneously after an electron is attached, however [10]. A previous study [5] revealed that the tautomerization of uracil was more accessible through intermolecular proton exchange (e.g., uracil↔glycine) than through the intramolecular PT of an isolated uracil. Thus, the dimer UG is employed here as an original model for investigations of one of the potential point mutations of DNA/RNA bases induced through interplay with amino acids. There are two different states of these biologically important anionic base molecules. One is the dipole-bound anionic state, such as the anionic base, which has been confirmed theoretically [11] and experimentally [12]. Another is the covalent anionic state, which

is unstable relative to the dipole-bound state [13], but can be stabilized by interacting with an amino acid. Interestingly, interaction with an amino acid leads to intermolecular barrier-free proton transfer (BFPT) from the O5 site of the acid to the O13 of the base (uracil or thymine) [10, 14, 15]. The hydrogenated bases that may result from intermolecular PT are predicted to play an important role in the mechanism of DNA and RNA damage [16].

Contrary to interaction with an amino acid, interaction with an alcohol can inhibit the BFPT of the base [16], which suggests a good way of regulating the PT or tautomerization of the base. How about methylating at either the base or amino acid moieties or both moieties of the dimer complex? The uracil in Fig. 1a is the most stable tautomer in the gas and solid phases as well as in solution [17, 18]. Methyl substitution at C11 can turn the uracil into thymine. After further methyl substitution at N9, the thymine will form 1-methylthymine (mT). Both anhydrous thymine [19, 20] and mT [21] have been studied experimentally with sophisticated techniques. Uracil in RNA and thymine in DNA are attached to the sugar at their N9 sites. Thus, this N9 site cannot participate in the tautomeric process in nucleotides of uracil and thymine. Generally, a damaged nucleotide can easily be excised by DNA/RNA polymerases if it is incorporated into a growing DNA/RNA strand, whereas strong hydrogen bonding with active-site residues can stabilize a uracil anion that has been removed from DNA via hydrolysis of the C–N9 glycosidic bond [22, 23]. Thus, the change in the C–N9 bond that occurs under physiological conditions can be observed if we use mT as a model to interact with other cellular components. On the other hand, the canonical glycine conformer in Fig. 1a was confirmed to be the most stable conformer both experimentally [24, 25] and theoretically [26, 27]. The principal structure of an amino acid $N_1RR–C_2RR–COOH$ involves a carbon backbone with a carboxyl group (COOH), an amino group (N_1RR), and an organic side group (R) of some kind that is linked to a central, tetrahedrally coordinated carbon atom (C_2). The side groups R of the amino acid in Fig. 1a can all be H (which corresponds to glycine), or (1) one R =

CH₃ at the C2 site and all the other R = H (corresponding to alanine), or (2) two R = CH₃ at the C2 site and all the other R = H (α-aminoisobutyric acid, mA), or (3) one R = CH₃ at the N1 site and all the other R = H (sarcosine, S), or (4) two R = CH₃ at the N1 site and all the other R = H (*N,N*-dimethylglycine, dmG) [28]. Studies of the interactions of glycine (G), alanine (A) and mA with base can partly show the effect of the side chain of the amino acid on the PT. Furthermore, the interaction of sarcosine or dmG with base can partly account for the enhancing effect of an amino acid or a peptide chain on PT.

Thus, in this paper, we will first discuss these neutral dimers. They are denoted U-G/A/dmA/S/dmG (U series), T-G/A/dmA/S/dmG (T series), and mT-G/A/dmA/S/dmG (mT series), and the effect of the methylation of the amino acid moiety on PT can be observed for them. We also define G series (G-U/T/mT), A series (A-U/T/mT),…, and dmG series(dmG-U/T/mT) dimers, in which the effect of methyl-substituted sequences and sites at the moiety of the base can be observed. Then we make a comparative study of the effect of methylation on anionic and cationized complexes due to the biological significance of low-energy electronic attachments [29, 30] and excitations [31, 32]. Maul et al. [32] revealed that electron excitation could induce the dissociation of the C2–C3 bond of an alanine molecule, a small deformation of glycine, and intramolecular PT for cysteine. Moreover, the positive charge can readily migrate from the excitation site along the peptide chain to the selected site, probably leading to C2–C3 bond breaking or isomerization [31]. Thus, the aim of present work was to probe not only the mutation and stability of bases, as regulated by amino acid interactions in neutral and anionic states, but also the counteracting actions of different bases on amino acids in the positive charged state. We hoped that our work would lead to deep insights into how charge migration in amino acid (or peptide) chains is regulated by base in DNA/RNA base–amino acid interaction pairs.

Computational methods

It is expected that the B3LYP and MP2 methods would give very similar results for the geometrical and vibrational features of bases [33], and DFT is an excellent compromise between computational cost and reasonable results. Therefore, in this article, B3LYP in combination with a considerably large basis set [6-311++G(d,p)] is adopted fully. The computed stationary points were characterized as minima or transition states by diagonalizing the Hessian matrix and analyzing the vibrational normal modes. In this way, the stationary points can be classified as minima if no imaginary frequencies are observed, or as transition states if only one imaginary frequency is obtained [34, 35].

After geometry optimization and frequency calculations (without scaling), zero-point energies and the sum of the electronic and thermal free energies can be obtained. All computed values are zero-point corrected energies (ZPCE) unless otherwise noted. Net atomic charges were obtained using the natural bond orbital analysis (NBO) of Weinhold et al. [36]. Analyses of the charge/spin distributions and molecular orbital information were also performed. All calculations were performed with the Gaussian 03 [37] software suite.

Results and discussion

Neutral complexes

Figure S1 in the "Electronic supplementary material" (ESM) displays the geometric structures, selected parameters and relative energies of the UG complex and fourteen methyl-substituted derivatives of it (including transition states as well as the anionic PT products). A series of PT processes can also be observed in the figure. Only the activation energies of these PT processes are listed in Table 1. Results reveal that only double-PT processes are available in the neutral UG complex and its methyl-substituted derivatives.

Effect of methyl substitution in uracil on double-PT processes

G series

Double PT in the UG dimer favors the isomerization of the uracil species due to its lower activation energy. For example, the most stable uracil U can transform into the third most stable isomer uracil U3 by exchanging protons H19 (N7 site of uracil) and H18 (O5 site of glycine) when interacting with an amino acid (see Fig. 1). Table 1 shows that the activation energies of the interconversions are only 10.3 (U→U3) and 1.5 (U3→U) kcal mol⁻¹, respectively. The isomerization of the most stable uracil can also be achieved by transferring H19 from its N7 site to its O13 site; i.e., by intramolecular PT (see details in [5]). The intramolecular PT process, however, costs as much as 41.7 kcal mol⁻¹ in activation energy, and the reverse process (U3 → U) also requires 29.4 kcal mol⁻¹ [5]. It is quite obvious that such a high energy disfavors the interconversion of the two equilibrium structures within the reproduction periods of typical biological species. However, this high activation energy can be greatly reduced by interacting with an amino acid [3, 4], which indicates the importance of the interaction for stabilizing the base isomers and establishing biomaterials such as RNA, cellulose [38], and other carbohydrates [39].

Table 1 Activation energies (ΔE) of a series of PT processes (kcal mol^{-1})

Neutral PT process	ΔE	Anionic PT process	ΔE	Cationic PT process	ΔE
UG→UG-2H	10.3	ani-UG→ani-UG-1H	No	cat-UG→cat-UG-1H	No
TG→TG-2H	9.2	ani-TG→ani-TG-1H	No	cat-TG→cat-TG-1H	0.5
mTG→mTG-2H	9.1	ani-mTG→ani-mTG-1H	No	cat-mTG→cat-mTG-1H	3.8
UA→UA-2H	8.8	ani-UA→ani-UA-1H	No	cat-UA→cat-UA-1H	No
TA→TA-2H	9.0	ani-TA→ani-TA-1H	No	cat-TA→cat-TA-1H	−1.6(0)
mTA→mTA-2H	8.8	ani-mTA→ani-mTA-1H	No	cat-mTA→cat-mTA-1H	−1.9(0.2)
UmA→UmA-2H	8.7	ani-UmA→ani-UmA-1H	No	cat-UmA→cat-UmA-1H	No
TmA→TmA-2H	9.0	ani-TmA→ani-TmA-1H	No	cat-TmA→cat-TmA-1H	0.6
mTmA→mTmA-2H	8.8	ani-mTmA→ani-mTmA-1H	No	cat-mTmA→cat-mTmA-1H	−1.5(0)
US→US-2H	8.9	ani-US→ani-US-1H	No	cat-US→cat-US-1H	−1.8(0.1)
TS→TS-2H	9.1	ani-TS→ani-TS-1H	No	cat-TS→cat-TS-1H	1.5
mTS→mTS-2H	9.0	ani-mTS→ani-mTS-1H	No	cat-mTS→cat-mTS-1H	−1.9(0)
UdmG→UdmG-2H	8.9	ani-UdmG→ani-UdmG-1H	No	cat-UdmG→cat-UdmG-1H	−1.2(1.2)
TdmG→TdmG-2H	9.2	ani-TdmG→ani-TdmG-1H	No	cat-TdmG→cat-TdmG-1H	−1.6(0)
mTdmG→mTdmG-2H	9.0	ani-mTdmG→ani-mTdmG-1H	No	cat-mTdmG→cat-mTdmG-1H	No

Data in parentheses are activation energies without ZPVE corrections

The activation energy of PT in the TG complex is 1.1 kcal mol^{-1} less than that in the UG complex, indicating that methylation at the C11 site of uracil is favorable to PT in the TG dimer and thus to the isomerization of the T species. In fact, the potential energy surfaces (PESs) of PT shown in Fig. 2 imply that the product TG-2H will convert to the reactant TG instantaneously, even if PT occurs. This is confirmed by the lower energy of the transition state compared to the product (0.4 kcal mol^{-1}) after including ZPCE (compare the PES in Fig. 2a with that in b). As reported by Pople et al. [40], the corresponding double PT does not exist in such processes, and the energy of the transition state is very sensitive to and extremely dependent on the ZPCE. Such ZPCE-dependent cases have also been observed in the PT processes of clusters formed by protonated bases (cytosine, uracil, thymine, and adenine) and ammonia [41].

As an extension, we further investigate the PT of a dimethylated product of uracil, mT, which is induced by interaction with glycine. Calculations reveal that further methylation at the N9 site of uracil decreases the PT activation energy by 0.1 kcal mol^{-1}. This indicates a more favorable PT process, although the decrease in energy is very small, and may even arise from calculational errors relating to B3LYP. The energy of the PT product mTG-2H is also lower than that of both the reactant mTG and the transition state, implying that the PT product is more stable. In the transition state of mTG → mTG-2H, O13–H18 (1.086 Å) and N7–H10 (1.270 Å) are shorter than the corresponding bonds of UG → UG-2H and TG → TG-2H, which indicates that mT (a dimethylated product of uracil) has a greater ability to attract two protons. As a result, we obtain a stable product (mTG-2H) after PT of the mTG. More importantly, the result reveals the different effects of methyl substitution at different sites in uracil on PT. We can also keep the product TG-2H stable by performing further methylation at the N9 site (TG-2H→mTG-2H).

Other amino acid series

For the A, mA, S and dmG series, their activation energies are all far lower than that of the intramolecular PT process for U↔U3, and they are also lower than those of the corresponding G series. A comparison reveals, however, that only the first PT process of each series (i.e., UA→UA-

Fig. 2a–b PESs with ZPCE (**a**) and without ZPCE (**b**) for the double-PT process of TG at 298 K

2H for the A series, UmA→UmA-2H for the mA series, US→US-2H for the S series, and UdmG→UdmG-2H for the dmG series) is available. The four reactants in the four PT processes actually belong to the U series and will be discussed in detail in the following section. The PT processes in the other amino acid series are inaccessible due to the equivalent or even higher energies of the PT products relative to those of the corresponding transition states.

Effect of methyl substitution in glycine on double-PT processes

After methylation at the C2 site of glycine, the glycine becomes alanine, and then UG becomes UA. Correspondingly, the G series becomes the A series. Table 1 shows that the activation energy in the UA→UA-2H process is lower by 1.5 kcal mol^{-1} than that in the UG → UG-2H one, indicating that methylation at the C2 site of glycine favors double PT. Dimethylating at the C2 site of the glycine causes UG to become UmA and the activation energy for PT to fall further, to 0.1 kcal mol^{-1}. This reveals that methylation of the C2 site of glycine favors DNA base isomerization.

Methylation at the N1 site of glycine can turn glycine into sarcosine. Accordingly, the activation energy for PT in US→US-2H is 1.4 kcal mol^{-1} lower than that in UG→UG-2H. Dimethylation at the N1 of G, causes glycine to turn into dmG. A comparison shows that the activation energy for PT in the UdmG→UdmG-2H process is equivalent to that (8.9 kcal mol^{-1}) for US→US-2H. This energy is higher than those for the UA→UA-2H and UmA→UmA-2H processes, implying that methylation at the C2 site instead of N1 of the glycine is more favorable to PT.

In a word, there are realizable UX →UX-2H processes (X = G, A, mA, S, dmG) for all of these UX complexes. This implies that the process U↔ U3 can be accomplished and regulated at lower energy by interacting with a different amino acid X. By contrast, methylation at either the C2 site or the N1 site of glycine can facilitate isomerization. In addition to the PT processes of the U series, only the mTG → mTG-2H and mTS→ mTS-2H processes in the other base series have stable PT products. Thus, the isomerization of uracil or thymine can be carried out or regulated by methylating at different sites and altering the number of methylations of the amino acid.

Anionic complexes

Effect of methyl substitution in uracil and/or glycine on geometries and PESs

Anionic UG (ani-UG) and its methyl-substituted derivatives have two features in common: (1) the unpaired electron localizes on a π* orbital of the base moiety (Fig. S2 of the ESM), similar to the valence anion of the isolated uracil [10, 14]; (2) there are BFPTs where one proton transfers from the carboxylic group of the amino acid moiety to the O13 atom of the base moiety (Fig. S1 of the ESM). The base moiety and the amino acid moiety bind much more strongly to each other after BFPT, as verified by the shortened hydrogen bonds between them. The driving force for the PT is the excess electron [14], which primarily localizes in the C10–C11–C12–O13 region of the uracil moiety. This phenomenon is consistent with earlier reports that the O13 site of uracil is susceptible to intermolecular PT in anionic complexes involving a uracil molecule and a weak acid [10, 14, 15].

The effects of methylating the uracil and/or glycine moiety on these BFPT processes are not striking (see Fig. 3). Figure 3a shows the adiabatic PES of the PT process (i.e., H18 → O13) for each of these fifteen complexes, and Fig. 3b presents the energy difference between the saddle-point structure and the corresponding anionic PT product shown in Fig. S1 of the ESM. The start length of the O5–H18 bond in Fig. 3a is 0.85 Å. The O13–H18 distance becomes 0.85 Å after PT. Clearly, all of these PT processes are energy free and thermodynamically favorable. Figure 3b shows that the differences for the G series are 3.3 (TG), 3.5 (UG), and 4.1 kcal mol^{-1} (mTG), respectively; whereas the corresponding differences for the A series are 3.8 (TA), 3.9 (UA), and 4.7 kcal mol^{-1} (mTA). Comparisons reveal that the differences for the other amino acid series lie between these two extremes. These data indicate that methylation at the C2 site of glycine is the one that is most energetically unfavorable to BFPT, although PT of dimers of the anionic A series is still spontaneous. On the other hand, the mT series among the base series have the most potential to inhibit BFPT, whereas the T series has the least. The U series lies between these two extremes. On the whole, the nature of the BFPT is unchanged, no matter which site or how many sites are methylated at either or both moieties of the ani-UG dimer.

Vertical detachment energies (VDEs) of these BFPT products

Due to the extra stabilization for the excess electron provided by the transferred proton, the VDE values of ani-UG-H and its methyl-substituted derivatives are about 28.0 kcal mol^{-1} higher than that of the isolated anionic uracil. Moreover, the electron is mainly found on the base moiety (Fig. S2 of the ESM). This indicates that the excess electron could have a stronger affinity for the protonated base species, and favors the stabilization of the dimer. This is why proton H18 of the amino acid species transfers to the O13 site of the base moiety spontaneously. The VDEs of these PT products are in the range 42.7~49.4 kcal mol^{-1}, which agrees well with the value (44.5 kcal mol^{-1}) for ani-
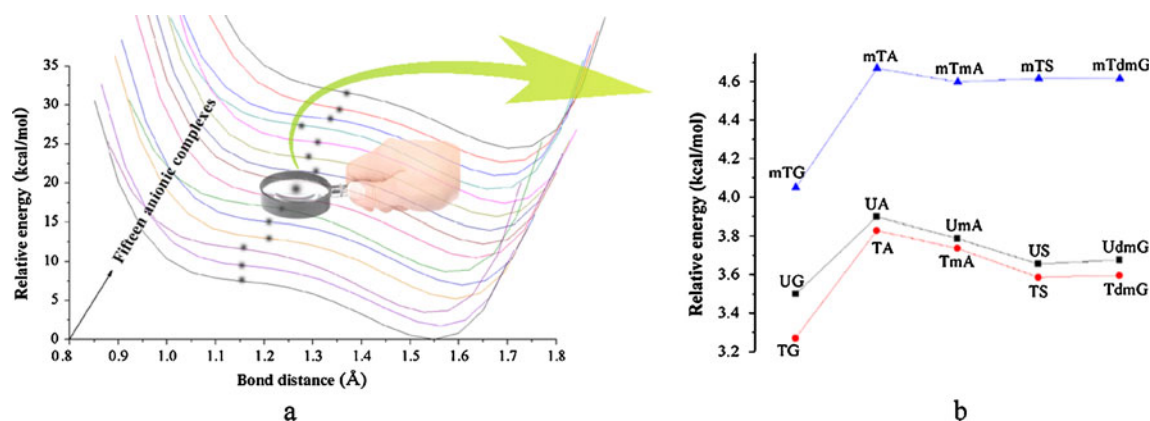
PT processes in **a** involve the transfer of H18 from the carboxylic group of the amino acid moiety to the O13 atom of the base moiety

UG-H reported by Gutowski et al. [10]. Generally, the values in Table S1 of the ESM systematically decrease as the number of methyl substitutions in the base moiety increases. These larger VDE values indicate that the base species will bind H18 strongly.

Note that not every mode of the anionic uracil–glycine dimer undergoes BFPT upon the attachment of an excess electron. For example, the neutral UG′ in Fig. 4 is the third most stable among its isomers [4]. Its anionic dimer ani-UG′ is linked by H19…O6 and O14…H18 bonds. Obviously, the initial H18 of the glycine does not transfer to the uracil moiety spontaneously to form ani-UG′-1H when the anionic dimer ani-UG′ is formed. Thus, we can predict that in this combined anionic structure, some activation energy is required if H18 is to transfer to O14. A similar phenomenon in which the electron was also located at the uracil moiety was also observed in anionic UG1 complexes by Gutowski et al. [10]. In combination with the HOMO results for the BFPT products in Fig. S2 of the ESM, we can expect that the extra electron in each of these anionic complexes will always position itself on the base moiety whether PT

happens or not. From the point of view of VDE, the VDE of ani-UG′ (29.4 kcal mol^{-1}) is about half of that of ani-UG-H (46.4 kcal mol^{-1}), indicating that electron detachment in the latter structure is far harder to achieve than in the former. Figure 4 shows that the excess electron still positions itself on the uracil moiety of ani-UG′. Thus, less VDE also implies that the excess electron in the ani-UG′ dimer attracts H18 of the glycine species rather weakly, disfavoring PT. This is why the ani-UG′ mode cannot transfer the proton spontaneously.

Cationic complexes

Effect of methyl substitution in uracil and/or glycine on geometries and the driving force for BFPT

Effect of methyl substitution in uracil and/or glycine on the length of the C2–C3 bond

BFPT occurs in most of the cationic complexes, just as it does in their anionic counterparts. In these cationic complexes, H18 departs from O5 and transfers to O13.
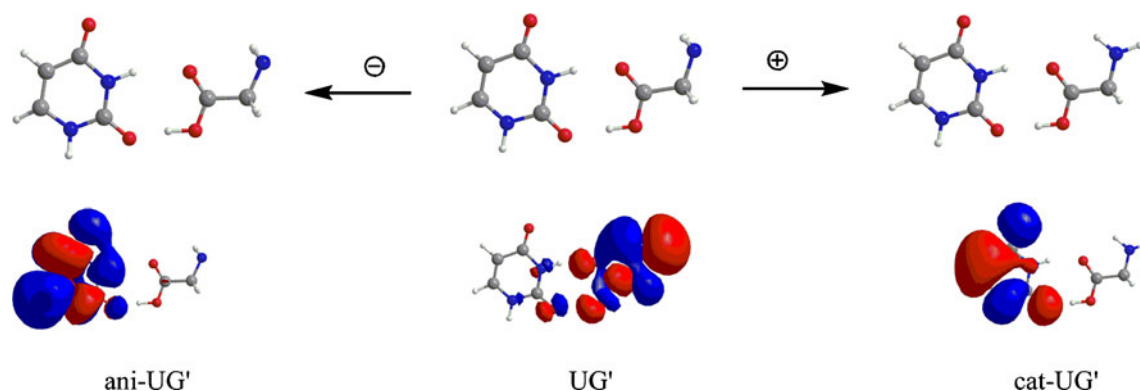

Fig. 4 HOMOs of UG′, which does not undergo BFPT upon the attachment of a charge

The C2–C3 bond lengths in all of the cationic complexes (see Fig. S3 of the ESM and Fig. 5) are extended by 0.093–0.543 Å compared to their corresponding neutral counterparts. The length of the C2–C3 bond also increases as the number of methyl substitutions in the uracil moiety increases. Methylation at the C2 site of the alanine moiety can also lengthen the bond. Multi-methylation at both moieties of the cat-UG dimer sometimes results in a conflicting effect. For example, the C2–C3 distance [i.e., 2.087 Å (cat-UmA-1H) > 2.045 Å (cat-TmA-1H, in Fig. 5) > 2.004 Å (cat-mTmA-1H)] decreases as the number of methyl substitutions in the base moiety increases. In fact, the bonds in cat-UmA-1H, cat-TmA-1H, and cat-mTmA-1H are broken, as confirmed by not only the long bond lengths (>2.0 Å) but also the positive charge distributions on both of the adjacent atoms. The bond lengths in TmA and cat-TmA are almost the same (1.539 vs 1.544 Å), however, suggesting that PT plays a key role in the dissociation of the bond. For complexes in the cat-G, cat-A and cat-S series, the C2–C3 bond in the PT product is also extended, but it is not broken, whereas the bond is broken in the isolated cationic alanine [32]. This result indicates that the interaction with the base can partially inhibit the dissociation of the C2–C3 bond. Only dimethylation at the C2 site of glycine (i.e., the cat-mA series) can induce C2–C3 dissociation.

Classes of BFPT processes

The PT products of the cationic complexes with BFPT in Fig. S3 of the ESM have similar structures to their anionic counterparts, as H18 transfers from the amino acid moiety to the base in all of these anionic and cationic complexes. The charge distributions in all of these cationic products is mainly concentrated on the base moiety (0.813~0.985), whereas the HOMO and the remaining unpaired electron are mainly located on the amino acid moiety (0.979~0.997). We can further subdivide these BFPT products into two classes, (a) direct BFPT products and (b) ZPCE-dependent ones. Obviously, the products of class (a) are cat-UG-1H, cat-UA-1H, cat-UmA-1H, and cat-mTdmG-1H, which can be obtained directly by optimizing their reactants, as done for the anionic complexes. The complexes in class (b), however, are sensitive to their ZPCE, so their PT processes are complicated and need to be explored in detail.

Features of the BFPT processes in class (b)

There are seven BFPT processes in class (b). In the process cat-TA → cat-TA-1H, the reactant can be turned into the product without needing to surmount an energy barrier, due to the equivalent energy of cat-TA and its transition state. The energy of cat-TA-H is 1.2 kcal mol^{-1} lower than those of both cat-TA and its transition state. Including ZPCE, the

transition state is found to be the most stable in energy (−1.6 kcal mol^{-1}), indicating that it is more favorable to the stability of the system to have H18 resonating between the base and amino acid species of the transition state. Likewise, the transition states in the cat-US → cat-US-1H, cat-UdmG → cat-UdmG-1H and cat-TdmG → cat-TdmG-1H processes are more stable than both the reactant and the product after the ZPCE has been included. Relative energy comparisons show that these four processes are characterized by similar stabilities of the reactant and the PT product. From the perspective of charge and electron distributions, we find that the first process is a PT one accompanied by charge and electron exchange. It can also be confirmed that the electron spin (0.374 vs. 0.626) and charge (0.402 vs. 0.598) distributions on the base moiety hardly differ from those on the amino acid in cat-TA, whereas the spin and charge distributions are each located almost entirely on different moieties after PT: 98.5% of the charge is on the base, while 98.9% of the electron spin is on the amino acid moiety in cat-TA-1H, respectively. Meanwhile, the two distributions in the transition state are almost intermediate between those in the reactant and the product, and H18 in the transition state carries a charge of +0.668. By comparison, in another three processes, the unpaired electron always positions itself on the amino acid moiety, and only the location of the charge distribution changes before and after PT, indicating these are three "pure PT" processes that do not involve electron transfer. From a structural point of view, the H-bond lengths in the transition states of these three processes [O13-H18 (1.222, 1.139, 1.269 Å) and H18-O5 (1.197, 1.296, 1.160 Å)] are shorter and very similar to each other, implying that H18 is the bridge in the base…amino acid resonance structures of the three transition states. Obviously, the resonance structure favors the greater stability of the corresponding transition state structure. In the first process, the lengths of the two hydrogen bonds are 1.424 and 1.070 Å; very different from each other.

Besides the four BFPT processes with lower-energy transition states mentioned above, there are another three BFPT processes in class (b): cat-mTA → cat-mTA-1H, cat-mTmA → cat-mTmA-1H, and cat-mTS → cat-mTS-1H. One feature that these three processes have in common is that their products are all several kcal mol^{-1} in energy lower than their corresponding reactants and transition states if the ZPCE is included. Moreover, their transition states (at the saddle point on the PES) are lower in energy than their corresponding reactants. Without ZPCE, all three reactants have the same energies as their transition states. Part of the unpaired electron on the base moiety and part of the charge on the amino acid in each reactant will transfer to the amino acid moiety (spin>90%) and the base (charge>90%) in the PT product, respectively, which indicates that both PT and electron transfer are involved in the last three reaction

processes. As well as the driving force from the stronger stability of the PT product, the electron transfer makes a favorable contribution to the three BFPT processes.

Effect of methyl substitution in uracil and/or glycine on the BFPT processes

Methylation at the base moiety will decrease the energy of both the transition state and the PT product relative to the nonmethylated ones. For example, the energies of cat-mTA-1H and its transition state are -5.6 and -1.9 kcal mol^{-1}, respectively. Both are lower in energy than their counterparts, cat-TA-H and its transition state (ts). This phenomenon can also be observed by comparing cat-mTS-1H/ts and cat-US-1H/ts, and by comparing cat-UdmG-1H/ts and cat-TdmG-1H/ts. Methylation at different sites on the amino acid moiety will lead to different effects. For example, methylation at the C2 site reduces the energy of the PT product but increases that of the transition state (comparing the processes of cat-mTA → cat-mTA-1H and cat-mTmA → cat-mTmA-1H); methylation at the N1 site, however, can increase the energies both the product and the transition state (comparing cat-US → cat-US-1H and cat-UdmG → cat-UdmG-1H). Thus, it is possible to regulate the reaction dynamics of these BFPT processes by methylating at different sites on a moiety or on a different moiety of the reactant. We claim that these BFPT processes are involved in proton transfer, not hydrogen atom transfer, due to two factors. One is that most of the positive charge (>0.9) is located at the base moiety in all of these PT products; the another is that about half of the positive charge is located on H18 in the transition state (see ts). For the anionic complexes, the excess electron is mainly located on the base species in the covalence-bound mode, just as it is for the isolated base [42]. The calculated charge distribution shows that the amino acid moiety possesses a charge of over -0.8. The H18 atom of the amino acid moiety transfers to the base moiety spontaneously. This implies that the transferred hydrogen is a single proton; this induces an anionic deprotonated amino acid moiety and a protonated base one. In this case, the base moiety still has a charge of about -0.16, indicating that the excess electron is still located at the base moiety. Thus, the single PT processes that occur in all of these anionic complexes are electron-assisted spontaneous ones.

PT processes requiring activation energy

Features of PT processes

Besides these BFPT cationic complexes, Fig. 5 displays another four dimers: cat-TG, cat-mTG, cat-TmA and cat-TS. These dimers require extra energy to facilitate proton

transfer. Their PT products are lower in energy than their corresponding reactants and transition states, demonstrating that the PT products are more stable than the reactants and are able to exist stably. Interestingly, the activation energies of these PT processes are not large ($0.5{\sim}3.8$ kcal mol^{-1}), indicating the kinetic feasibility of the PT. The HOMOs of the four reactant dimers are located on both the base and amino acid moieties, as mentioned above. The spin distribution over each moiety is in excellent agreement with the HOMO distribution. For example, the HOMO of cat-TS shows that the electron is mainly located on the S moiety, which is highly consistent with the spin distributions of both moieties (T: 0.120 vs. S: 0.880). The charge distributions show that most of the charge on these dimers (except for cat-mTG) is located on the AA moiety, however. Once the PT products have formed, the charge will redistribute, and almost all of the charge congregates on the amino acid moiety. Cat-mTG is the only dimer that has more charge on the base moiety (0.510) than on the amino acid one (0.490), highlighting the potential difficulty involved in achieving PT from the amino acid moiety to the base one. The largest activation energy of 3.8 kcal mol^{-1} in the cat-mTG→cat-mTG-1H process also suggests this difficulty. By contrast, the activation energies in the cat-TG→cat-TG-1H and cat-TmA→cat-TmA-1H processes are lower and almost identical ($0.5{\sim}0.6$ kcal mol^{-1}), indicating two accessible PT reactions. The activation energy in the cat-TS→cat-TS-1H process is 1.5 kcal mol^{-1}, larger than that in the above two processes. The reactants in the latter three processes differ in their amino acid moieties (glycine, mA and sarcosine). Thus, their different activation energies lead to two different methylation effects: methylation at the C2 site will hardly affect the kinetics of PT, but methylation at the N1 site will hinder PT. This observation is similar to that noted for the above BFPT processes.

These PT processes are concerted ones, with electron transfer from the base to the amino acid and PT (or charge transfer) from the amino acid to the base. For example, in the transition state of the cat-TG→cat-TG-1H process, the charge on the proton is 0.498 and the distance H18–O5 is 1.106 Å, indicating that H18 is still located at the AA moiety and holds a positive charge. Moreover, the HOMO (or electron spin) has moved partly from the initial base moiety to the amino acid. In the product of cat-TG-1H, the base moiety holds almost a whole unit of charge (0.894) and the H18–O5 distance has also extended to 1.644 Å.
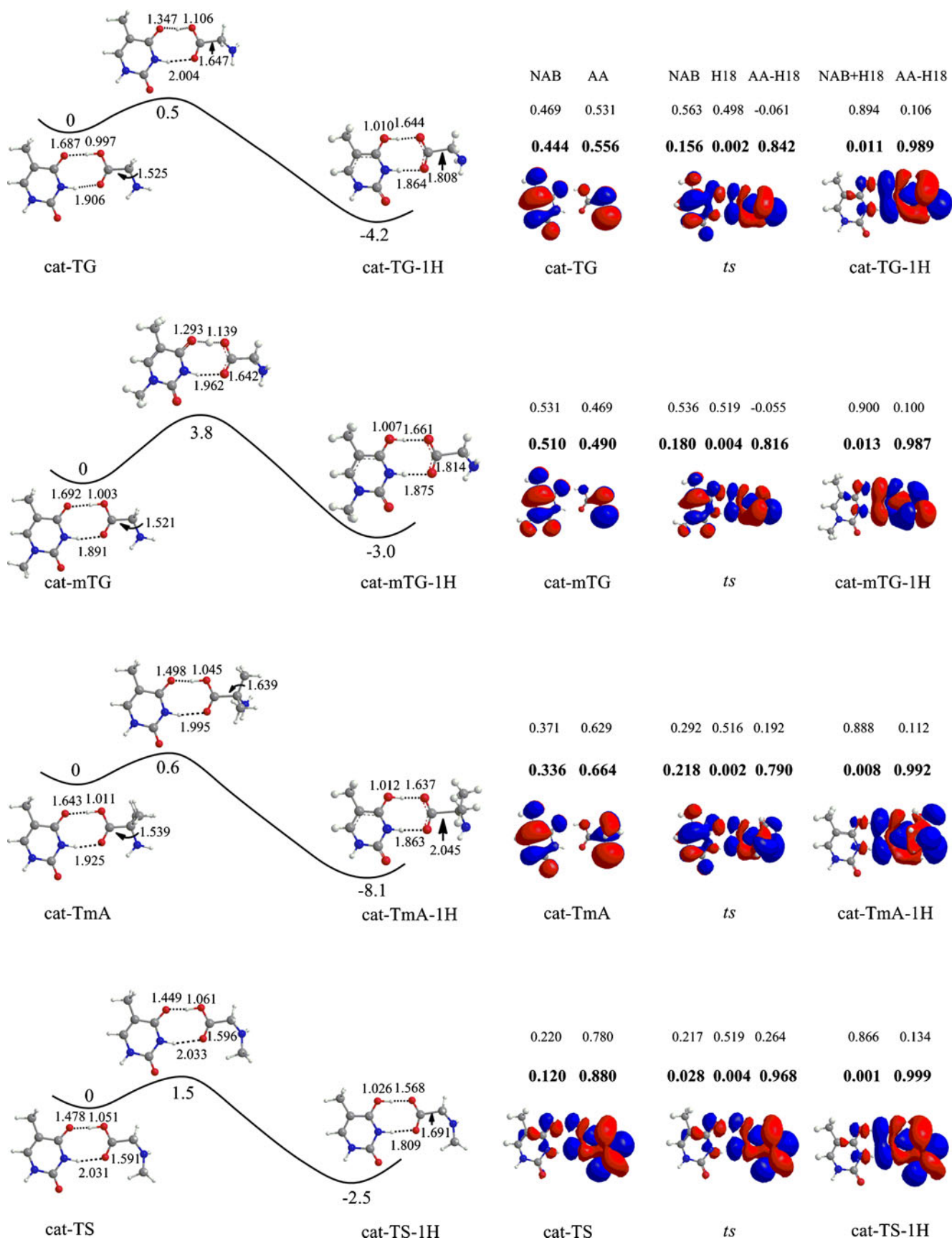
Meanwhile, the HOMO is almost completely located on the amino acid moiety (spin: 0.989). It is worth noting that the C2–C3 bonds in the reactant, transition state and product in the cat-TmA→cat-TmA-1H process are 1.539, 1.639 and 2.045 Å in length, respectively, indicating that bond dissociation is slower and governed by PT and charge migration [31]. It is interesting that the cat-TmA reaction has a much lower minimum (to cat-TmA-1H), −8.1 kcal mol^{-1}, than the other three reactions (which are in the range of −2.5 to −4.2). Analysis shows that cat-TmA is the only dimer with like charges on its C2 (−0.109) and C3 (−0.172) atoms. Thus, the poorer stability of cat-TmA derives from the stronger like-charge repulsion between the C2 and C3 atoms. On the other hand, an extended or dissociated C2–C3 bond will enhance the stability of the complex. This is why the PT product cat-TmA-1H has greater stability and a longer C2–C3 bond (2.045 Å) than its reactant cat-TmA (1.539 Å).

Directions of charge migration in the four cationic dimers that require activation energy for PT

It can be seen from Fig. 5 that the charge is mainly located on the amino acid moiety (>0.5) in the three reactants cat-TG, cat-TS, and cat-TmA before PT. However, the charge in cat-mTG is mainly distributed on the base moiety (0.510). Once PT has occurred, most of the positive charge in these PT products migrates to the base moiety. This charge greatly facilitates the PT, as all of the PT processes in the neutral complexes require significant activation energies, whereas those in the cationic ones require less or even no activation energy. In a peptide, the charge prefers to move from the ring group (which is electron rich) at one end of the peptide to the N-terminus at another end [31]. In dimer systems with intermolecular PT, it also easily migrates, but most of the charge moves from the amino acid moiety to the base moiety. This is just the opposite to the direction of charge migration in a peptide [31]. This difference may be due to the exchange of positive charge and an electron between the two moieties in the cationic dimers. Identical directions of charge transfer can be observed for the BFPT processes in Fig. S3 of the ESM.

Switch effect

Recently, an interesting observation was reported by Bickelhaupt's group [43]. They noted that the hydrogen bonds between guanine and cytosine would deform and act as a supramolecular switch (the "switch effect") when an anion or cation was substituted onto the guanine or cytosine in the GC pair. The switch effect was also an important factor for inducing a point mutation. Thus, in the present work, we also probed this type of switch effect in the

neutral and charged UG and its methylated derivatives (see Fig. 1b and c).

Obviously, the difference (Δr) between r_1 (the O13… H18–O5 distance) and r_2 (the N7–H19…O6 distance) can be taken as an approximate parameter for estimating the degree of switching (Table 2). The Δr values of neutral complexes with a methylated amino acid moiety are less than those of complexes with a methylated base moiety, so the switch effect is less prominent in the former type of complex. In the charged complexes, Δr is obviously much larger, indicating a strong effect of charge on the switch effect. For anionic complexes, Δr increases with the number of methyl substitutions in the base moiety, whereas it decreases with the number of methyl substitutions in the amino acid moiety. These results suggest a marked switch effect in the above species. As expected, the switch effect in the cationic complexes shows different behavior. For example, the differences between r_1 and r_2 for the cat-U series are 0.236 (cat-UG-1H), 0.270 (cat-UA-1H), 0.247 (cat-UmA-1H), 0.267 (cat-US-1H) and 0.244 Å (cat-UdmG-1H), respectively; for the cat-G series they are 0.236 (cat-UG-1H), 0.244 (cat-TG-1H) and 0.240 Å (cat-mTG-1H), respectively. The results indicate that the methylation at either C2 or N1 on glycine can greatly open the switch. However, further methylation at one of the two sites will close the switch somewhat. Methylation at the base moiety also yields a similar effect.

Based on the above discussions, we can conclude that the charge plays an important role in the switch effect by opening or closing the r_1 and r_2 "mouth" of the UG dimer and its various methylated derivatives. Methylation mainly affects the configuration of the complexes to some degree.

Conclusions and remarks

A systematic study on the potential interactions between DNA and RNA-protein was performed, using simplified models for the UG dimer and its various methylated derivatives as well as their corresponding charged species. This study allowed the following important conclusions to be drawn.

Intermolecular proton exchange is enhanced if hydrogens are replaced with methyls in the base moiety of each neutral complex in the amino acid series. This is also true if hydrogens are replaced with methyls in the amino acid moiety of each complex in the base series. These results imply that the induction of a point mutation by PT in the DNA/protein interaction would be more feasible than in the RNA/protein interaction. When ZPCE is not included, the order of PT activation energy is: T series > U series > mT series. When ZPCE is included, the activation energies of the corresponding dimers in the U series and the mT series

Table 2 Hydrogen bond lengths (Å)

Complexes	Neutral			Anionic			Cationic		
	r_1	r_2	Δr	r_1	r_2	Δr	r_1	r_2	Δr
UG-1/2H	2.686	2.867	0.181	2.541	2.831	0.290	2.648	2.884	0.236
UA-1/2H	2.711	2.876	0.165	2.579	2.846	0.267	2.612	2.892	0.280
UmA-1/2H	2.713	2.873	0.160	2.579	2.837	0.258	2.642	2.889	0.247
US-1/2H	2.710	2.876	0.166	2.581	2.848	0.267	2.581	2.848	0.267
UdmG-1/2H	2.716	2.876	0.160	2.579	2.844	0.265	2.557	2.819	0.262
TG-1/2H	2.684	2.873	0.189	2.538	2.837	0.299	2.653	2.897	0.244
TA-1/2H	2.708	2.880	0.172	2.579	2.847	0.268	2.644	2.894	0.250
TmA-1/2H	2.709	2.878	0.169	2.579	2.846	0.267	2.648	2.895	0.247
TS-1/2H	2.707	2.882	0.175	2.580	2.848	0.268	2.592	2.846	0.254
TdmG-1/2H	2.710	2.884	0.174	2.576	2.848	0.272	2.555	2.809	0.254
mTG-1/2H	2.683	2.890	0.207	2.537	2.851	0.314	2.666	2.906	0.240
mTA-1/2H	2.698	2.886	0.188	2.576	2.849	0.273	2.656	2.905	0.249
mTmA-1/2H	2.700	2.883	0.183	2.575	2.849	0.274	2.659	2.900	0.241
mTS-1/2H	2.697	2.888	0.191	2.577	2.849	0.272	2.607	2.857	0.250
mTdmG-1/2H	2.702	2.887	0.185	2.575	2.848	0.273	2.571	2.820	0.249

The lengths in all these complexes are for the PT products, since lengths in most of the corresponding charged reactants before PT are not available. Only the lengths in the neutral PT products are comparable to those in the charged products. Calculated at the B3LYP/6-31++G** level. $\Delta r = r_2 - r_1$

are almost identical. Above all, the energies of some of the neutral products of double PT are larger than those of the corresponding transition states when ZPCE is included, indicating that double PT is not available. A similar phenomenon was also observed in protonated base–ammonia complexes [41].

Attaching an extra electron can induce the BFPT of H18 from the carboxylic group of the amino acid moiety to the O13 atom of the base moiety (i.e., one-way PT). The PT will result in protonated and deprotonated species for the base and amino acid moieties, respectively. Thus, the driving force for the BFPT is the stabilization of the excess electron. This is a PT process with electron assistance. Methylation at the amino acid moiety will disfavor the BFPT, but it is insufficient to turn the BFPT into a PT process that requires energy. Thus, such methylation can be used to fine-tune the BFPT. In fact, it is the most effective method of inhibiting the BFPT completely to alter the mode of interaction of the base and amino acid; i.e., we can change the BFPT process into the a process that requires activation energy by changing the mode of interaction of the base and amino acid [10]. Moreover, methylation in a complex with such an interaction mode favors rather than inhibits proton exchange.

A positive charge can result in not only BFPT but also extended C2–C3 distances in the seven dimers compared to the case with electron attachment [32]. The driving force for this PT differs from that seen in anionic complexes. In anionic PT products, the extra electron is located at the base moiety, and the moiety is also a proton acceptor. In the seven cationic complexes, the amino acid moiety becomes an electron acceptor and a proton donor, whereas the base

moiety is the charge/proton acceptor. These BFPT processes can be divided into two classes: direct BFPT processes and ZPCE-dependent ones. For the latter, the energies of some transition states are lower than those of the corresponding reactants and PT products once the ZPCE is included. Analyses show that only the cat-UdmG → cat-UdmG-H and cat-TdmG → cat-TdmG-H processes are pure PT (or charge transfer) processes. The other five ZPCE-dependent processes are concerted ones, with both PT (or charge transfer) and electron transfer. In addition, there are four different methylations that require some activation energy to turn cat-UG into cat-UG-1H. These are: mono-methylation at the C11 site of the uracil (cat-TG), dimethylation at the C11 site of the glycine (cat-mTG), both mono-methylation at the C11 site of the uracil and dimethylation at the C2 site of the glycine (cat-TmA), and mono-methylation at the C11 site of the uracil and at the N1 site of the glycine (cat-TS).

In all of these cationic dimer systems, the original charge on the amino acid moiety easily migrates to the ring-structure base moiety, whereas the unpaired electron on the base moiety transfers to the amino acid moiety after PT. The direction of charge migration is opposite to that seen in a peptide [31]. The migration also induces the extension or even dissociation of the C2–C3 bond of the amino acid moiety, as it was found to in a peptide main chain [32]. Hence, we can regulate the magnitude of the activation energy of PT and the degree of extension of the C2–C3 bond in a DNA/RNA-protein system by altering the site(s) or number of methylations. Charge plays a more important role in determining whether the dimer switch is open or closed, whereas methylation mainly affects the configura-

tion of the dimers, and can be used to regulate the switch to a certain extent.

Although the model employed in the paper is simple, and represents only one of the potential modes of RNA/DNA-protein interaction, it still provides a lot of valuable information. For example, point mutations induced by the isomerization of uracil or thymine bases of RNA/DNA can be inhibited by binding charged amino acids, or point mutations can be intensified by binding uncharged amino acids, as discussed above. Other modes of base–amino acid interaction will be probed in the near future so that more comprehensive RNA/DNA-protein interactions can be simulated in simplified models.

References

1. Topal MD, Fresco JR (1976) Nature 263:285–293
2. Lowdin PO (1965) Adv Quantum Chem 2:213–360
3. Dabkowska I, Gutowski M, Rak J (2005) J Am Chem Soc 127:2238–2248
4. Dabkowska I, Rak J, Gutowski M (2002) J Phys Chem A 106:7423–7433
5. Li D, Ai H (2009) J Phys Chem B 113:11732–11742
6. Shaw RW (2003) Kinetic biochemistry and molecular biology. University of Florida, Gainesville
7. Shaw RW, Feller JA, Bloom LB (2004) DNA Repair 3:1273–1283
8. Luscombe NM, Laskowski RA, Thornton JM (2001) Nucleic Acids Res 29:2860–2874
9. Suzuki M (1994) Structure 2:317–326
10. Gutowski M, Dabkowska I, Rak J, Nilles JM, Bowen KH (2002) Eur J Phys D 20:431–439
11. Oyler NA, Adamowicz L (1993) J Phys Chem 97:11122–11128
12. Desfrancois C, Abdoul-Carime H, Schermann JP (1996) J Chem Phys 104:7792–7794
13. Bachorz RA, Rak J, Gutowski M (2005) Phys Chem Chem Phys 7:2116–2125
14. Haranczyk M, Bachorz R, Rak J et al (2003) J Phys Chem B 107:7889–7895
15. Radisic D, Bowen KH Jr et al (2005) J Am Chem Soc 127:6443–6450
16. Harańczyk M, Rak J, Gutowski M et al (2005) J Phys Chem B 109:13383–13391
17. Kryachko ES, Nguyen MT, Zeegers-Huyskens T (2001) J Phys Chem A 105:1288–1295
18. Tian SX, Zhang CF, Zhang ZJ, Chen XJ, Xu KZ (1999) Chem Phys 242:217–225
19. Bowers MT, Kemper PR, von Helden G, van Koppen P (1993) Science 260:1446–1451
20. Clemmer DE, Jarrold MF (1997) J Mass Spectrom 32:577–592
21. Wyttenbach T, von Helden G, Bowers MT (1996) J Am Chem Soc 118:8355–8364
22. Drohat AC, Xiao G, Tordova M et al (1999) Biochemistry 38:11876–11886
23. Werner RM, Stivers JT (2000) Biochemistry 39:14054–14064
24. Suenram RD, Lovas FJ (1978) J Mol Spectrosc 72:372–382
25. Lovas FJ, Kawashima Y, Grabov JU et al (1995) Astrophys J 455: L201–L204
26. Jensen JH, Gordon MS (1991) J Am Chem Soc 113:7917–7924
27. Csaszar AG (1992) J Am Chem Soc 114:9568–9575
28. Wyttenbach T, Witt M, Bowers MT (2000) J Am Chem Soc 122:3458–464
29. Desfrancois C, Carles S, Schermann JP (2000) Chem Rev 100:3943–3962
30. Wesolowski SS, Leininger ML, Pentchev PN, Schaefer HF (2001) J Am Chem Soc 123:4023–4028
31. Weinkauf R, Schlag EW, Martinez TJ, Levine RD (1997) J Phys Chem A 101:7702–7710
32. Maul R, Preuss M, Ortmann F, Hannewald K, Bechstedt F (2007) J Phys Chem A 111:4370–4377
33. Hu X, Li H, Liang W, Han S (2004) J Phys Chem B 108:12999–13007
34. Malick DK, Petersson GA, Montgomery JA Jr (1998) J Chem Phys 108:5704–5713
35. Peiró-García J, Nebot-Gil I (2003) Chem Phys Chem 4:843–847
36. Weinhold F, Carpenter JE (1988) Dissolution of cellose with ionic liquids. Plenum, New York
37. Frisch MJ, Trucks GW, Schlegel HB et al (2004) Gaussian 03, revision C.02. Gaussian, Inc., Wallingford
38. Swatloski RP, Spear SK, Holbrey JD, Rogers RD (2002) J Am Chem Soc 124:4974–4975
39. Forsyth SA, MacFarlane DR, Thomson RJ, von Itzstein M (2002) Chem Commun 714–715
40. Pople JA, Raghavachari K, Frisch MJ, Binkley JS, Schleyer PvR (1983) J Am Chem Soc 105:6389–6398
41. Wu RH, McMahon TB (2007) J Am Chem Soc 129:569–580
42. Sevilla MD, Besler B, Colson AO (1995) J Phys Chem 99:1060–1063
43. Guerra CF, Wijst T, Bickelhaupt FM (2006) Chem Eur J 12:3032–3042

ORIGINAL PAPER

Theoretical study on chiral recognition mechanism of methyl mandelate enantiomers on permethylated β-cyclodextrin

Jie-hua Shi · Zuo-jing Ding · Ying Hu

Abstract Host-guest interactions of permethylated β-cyclodextrin (PM-β-CD) with methyl mandelate enantiomers ((R/S)-MMA) were simulated using semiempirical PM3 and ONIOM (B3LYP/6-31G(d):PM3) method. The chiral recognition mechanism of (R/S)-MMA enantiomers on PM-β-CD was investigated. The binding energies for all orientations considered in this research are reported. The most stable geometry structures of the two complexes are different. The benzene ring of (R)-MMA locates horizontally approximately on the wider edge of the PM-β-CD cavity, but the aromatic ring of (S)-MMA is deeply included into the hydrophobic cavity. Furthermore, the results of NBO analysis show that the main driving forces in the inclusion process of PM-β-CD with (R/S)-MMA are hydrogen bonding interaction, dipole-dipole interaction, charge-transfer and hydrophobic interaction. The stabilization energy of the (R)-MMA/PM-β-CD complex is lower than that of the (S)-MMA/PM-β-CD complex. Moreover, the chiral carbon in MMA of (R/S)-MMA/PM-β-CD complexes are close to the C2 and C3 in the glucose unit. The chiral recognition mechanism is thus closely related to the chiral environment provided by C2 and C3 in the glucose unit and the degree of (R/S)-MMA and PM-β-CD inclusion.

J.-h. Shi (✉) · Z.-j. Ding · Y. Hu
College of Pharmaceutical Sciences,
Zhejiang University of Technology,
Hangzhou 310032, China
e-mail: shijh@zjut.edu.cn

J.-h. Shi
State Key Laboratory Breeding Base of Green Chemistry
Synthesis Technology, Zhejiang University of Technology,
Hangzhou 310032, China

Introduction

If the amylose fraction of starch is degraded by glucosyltransferases, one or several turns of the amylose helix are hydrolyzed off and their ends are joined together, thereby producing cyclic oligosaccharides called cyclodextrins (CDs) or cycloamyloses (CAs) [1]. Commonly, CDs refers to the cyclic oligosaccharides with 6, 7, and 8 residues named α-, β-, and γ-CD, respectively. CAs, also called large-ring cyclodextrins, have many more residues per cycle. CDs and CAs are both important in their own right because they can sequester guest molecules in their internal cavities and can function as adsorbents or time-release agents [2]. Generally, CDs molecules are more suitable to accommodate small guest molecules which fit spatially within the cavities formed by the annular structures [1]. The molecular form of CDs can be characterized as truncated cones with the secondary hydroxyl groups located at the wider edge of the ring and the primary hydroxyl groups on the narrower edge, so the inner side of the cavity is hydrophobic and the outer side is hydrophilic. Due to this unique chemical structure of CDs, they are easy to form inclusion complexes with a great variety of guest molecules without covalent bond. The resultant inclusion complexes frequently exhibit obvious differences in their physico-chemical properties of guest molecules (such as water solubility, the bioavailability, the stability of drug molecules, and so on) [3–5]. Furthermore, CDs can form diastereomeric complexes with chiral guest molecules

because of the chiral character of the D-glucose units. This property offers the possibility to utilize CDs-based materials in the field of the enantioselective separation of chiral compounds. They, as chiral recognition selectors, have been widely used in separation of enantiomers of chiral compounds [6–9]. However, the ability of native CDs to achieve enantioseparation of chiral compounds was poor due to the molecular rigidity of native CDs, whereas cyclodextrin derivatives with higher molecular flexibility, which hydroxyl groups in CDs were replaced with other groups, can improve their capability for chiral recognition because they allow a conformational adjustment to each enantiomer [10–12].

With the increasing number of applications of CDs and their derivatives in separation science, the chiral recognition mechanisms underlying the enantioselective separation in the presence of CDs and their derivatives as chiral selectors should provide insights. Thus, it is necessary to clarify the structures of the complexes and elucidate the separation mechanisms because the enantioseparation mainly results from important differences between the complexes geometries. Owing to the unique advantages, molecular modeling methods have been valuable tools for elucidating the chiral recognition mechanisms, which are frequently used to rationalize experimental findings concerning chiral recognition by CDs and their derivatives.

Currently, there are an increasing number of theoretical methods used in supramolecular systems. Among these methods, molecular mechanics (MM) [13], molecular dynamics (MD) [14] and semiempirical methods [15–17] (such as AM1, PM3, among others) are the most widely used in the relatively larger size and numerous atoms of CDs and their derivatives, as *ab initio* and density functional theory (DFT) methods are prohibitively expensive to carry out studies on such large systems. The accuracy of PM3 calculation is lower than *ab initio* and density functional theory (DFT) methods. To overcome these drawbacks, Morokuma et al. [18–20] developed a hybrid ONIOM (our Own N-layer integrated orbital molecular mechanics) method to perform a complex geometry optimization. The hybrid ONIOM method has captured many researchers' interest. It can treat different parts of a system simultaneously with good accuracy and lower computational cost compared to *ab initio* and DFT methods, and it has been proved to be effective for investigating inclusion complexes of the CDs or their derivatives with guests [21–25].

Mandelic acid and its derivatives are important chiral intermediates in asymmetric synthesis reactions, which are widely used to produce a variety of optically pure amino acids, angiotensin converting enzyme inhibitors, coenzyme A, among others [26]. If these hydroxy acid enantiomers can be separated effectively, it is significant for the study of asymmetric syntheses and the quality control of pharmaceu-

tical production process. The enantioseparation have been reported [9, 14, 26]. However, the mechanism for most separations is still unknown in detail, so it is of particular significance to get an insight into this topic. Lipkowitz et al. [14] have carried out a multifaceted study of methyl mandelate binding to β-CD using chromatography to derive thermodynamic data, and they have also used NMR spectroscopy to discern binding sites, and molecular simulations to investigate the recognition process at atomic level. However, to the best of our knowledge, the chiral recognition mechanism of methyl mandelate enantiomers on PM-β-CD by PM3 and ONIOM method are not yet reported.

In this study, the energy and geometry of PM-β-CD complexations with (R/S)-MMA enantiomers were studied by PM3 and ONIOM method. The aim of this study is to discover why and how the chiral recognition takes place, and to study the main interaction forces between PM-β-CD and (R/S)-MMA enantiomers. To study the enantiodifferentiation of (R/S)-MMA by PM-β-CD, we first determine the interaction energies between PM-β-CD and (R/S)-MMA enantiomers, then establish the differences in the corresponding inclusion complexes.

Computational methods

Molecular modeling and quantum chemical calculation

The starting structures of (R/S)-MMA were constructed using Chem 3D Ultra (Version 8.0, CambridgeSoft com, USA), and were fully optimized using PM3 method implemented in Gaussian 03 (Gaussian Inc., Wallingford, USA) until all eigenvalue of the Hessian matrix were positive [27]. The stable conformations of (R/S)-MMA were obtained with a minimum energy. The heats of formation of the optimized (R)-MMA and (S)-MMA are −438.28 and −438.49 kJ mol^{-1}, respectively.

The initial geometry of PM-β-CD was constructed, using Chem 3D Ultra, from the crystallographic parameters of β-CD taken from the Cambridge structural database (CSD). The H-atoms in hydroxyl groups of β-CD were replaced with methyl groups. The structure was energy minimized using PM3 method implemented in Gaussian 03 until all eigenvalue of the Hessian matrix were positive. And the heat of formation of the optimized PM-β-CD is −5520.48 kJ mol^{-1}. The PM3-optimized structures of PM-β-CD and (R/S)-MMA were used for the molecular docking calculations.

The coordinate system used to define the inclusion process of PM-β-CD with (R/S)-MMA is shown in Fig. 1. The construction method is reported in the literature [28]. Briefly, glycosidic oxygen atoms (O$_a$, O$_b$, O$_c$) of the optimized PM-β-CD are placed onto the *XY* plane and

their center is defined as the origin of the Cartesian coordination system. The Z-axis is perpendicular to the XY plane. The 2-OMe and 3-OMe groups in each glucose uint are placed pointing toward the positive Z-axis. The (R/S)-MMA molecules were docked into the cavity of PM-β-CD by four types: the first docking orientation is that the bond linking the chiral carbon to the benzene ring (C7-C5) is placed on the Z-axis and the carbonyl is head up orientation (Fig. 1 type 1); the second is opposite to the first as the carbonyl is head down orientation (Fig. 1 type 2); the third docking approach is that the bond linking the chiral carbon and the hydroxyl oxygen (C7-O11) is placed on the Z-axis and the carbonyl is head down orientation (Fig. 1 type 3); and the fourth is opposite to the third as the carbonyl is head up orientation (Fig. 1 type 4). The relative position between the host and the guest was measured by the Z-coordinate of the chiral carbon atom (C*) of the guest (Fig. 1). The inclusion process emulation was then achieved

Fig. 1 Docking procedure for (R/S)-MMA into PM-β-CD: (a) PM-β-CD construct, (b) (R/S)-MMA construct, (c) docking orientations
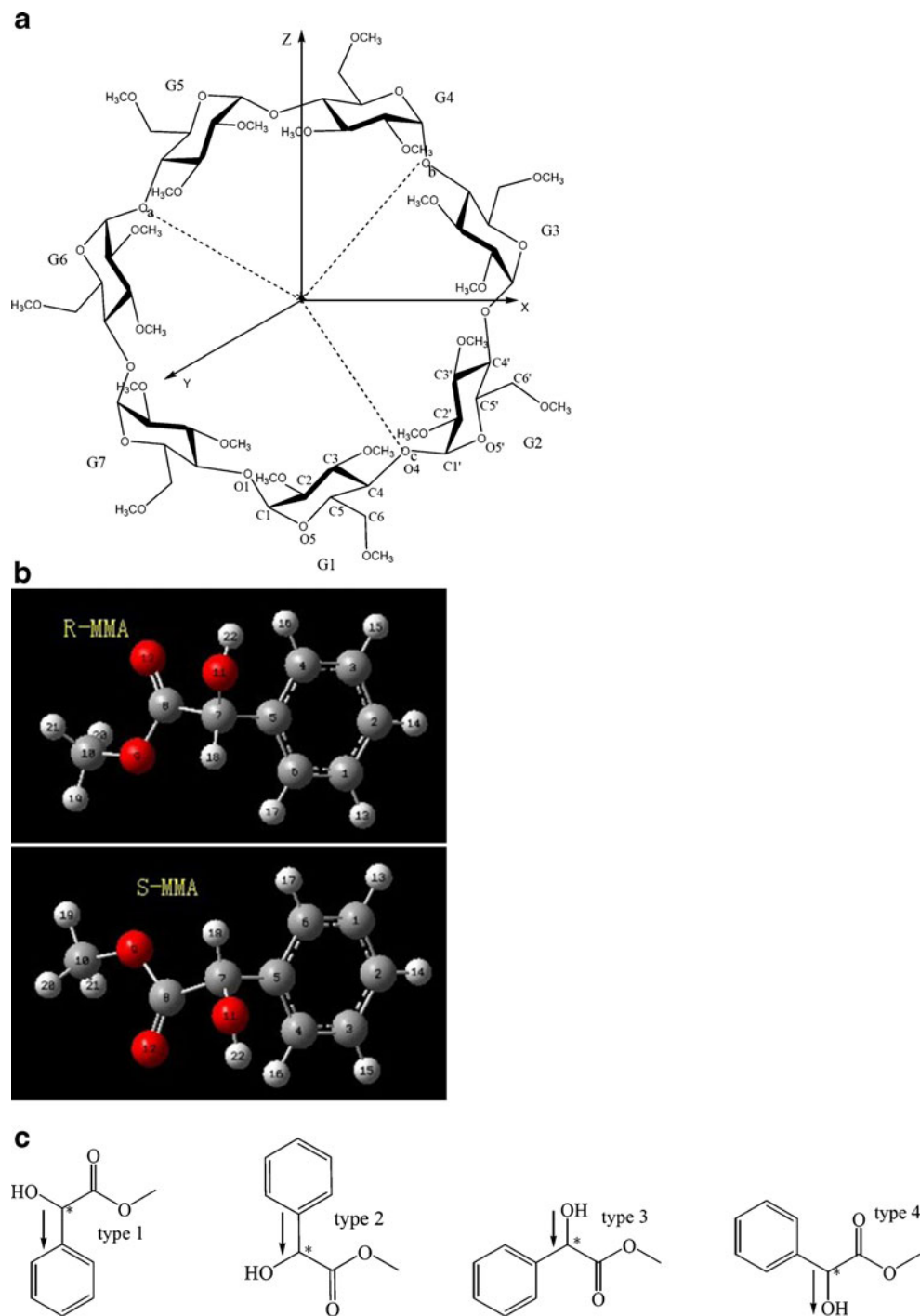
along the Z axis from 8 to −8 Å with a step of 1 Å, and a systematic search of the (R/S)-MMA molecules around Z-axis, from −180° to 180°, was performed using a grid of 30°. So, 1632 structures, which is all combinations of 17(translations)×12(rotations)×4(guest orientations)×2(enantiomers), have been calculated using PM3 method in this work. These complex geometries were energy minimized using PM3 method until all eigenvalues of the Hessian matrix were positive. All optimizations were carried out in vacuo. However, the statistical thermodynamic calculation was performed at 298.15 K at 1 atm, for the PM3-optimized equilibrium geometries of the (R/S)-MMA/PM-β-CD inclusion complexes.

The two most stable structures optimized by PM3 method for (R/S)-MMA/PM-β-CD complexes were further optimized using a two-layered hybrid ONIOM method. The high-level layer, (R/S)-MMA, and low-level layer, PM-β-CD, were treated by B3LYP/6-31G(d) level of theory and PM3 method, respectively.

Finally, natural bond orbital (NBO) calculation at B3LYP/6-31G(d) level, for the most stable (R/S)-MMA/PM-β-CD complex optimized by ONIOM method, was carried out in order to elucidate the intermolecular interaction between host and guest molecules via the determination of the stabilization energy $E^{(2)}$.

Definition of the binding energy (BE) and the total ONIOM energy

In order to investigate the driving forces leading to the actual complex between (R/S)-MMA and PM-β-CD, the binding energy (BE) upon complexation between host and guest calculated for the minimum energy structures is defined in Eq. (1) [17]:

$$BE = E(\text{host} - \text{guest})^{opt} - [E(\text{host})^{opt} + E(\text{guest})^{opt}] \quad (1)$$

where $E(\text{host-guest})^{opt}$, $E(\text{host})^{opt}$, and $E(\text{guest})^{opt}$ represent the heats of formation of the complex, the free PM-β-CD and the free guest, respectively.

The deformation energy for each component, the guest or the host molecule, throughout the formation of the complex, was defined as the difference in the energy of the totally optimized component compared to its energy in the complex as Eq. (2) [29]:

$$DEF(\text{component}) = E(\text{component})_{sp}^{opt} - E(\text{component})^{opt} \quad (2)$$

DEF (component) stands for deformation energy of the guest or the host. $E(\text{component})_{sp}^{opt}$ is the single point energy of the component on the configuration taken from the optimized complex geometry. The component is defined as host or guest.

In the two-layered ONIOM method, the molecular system is divided into an inner and an outer layer. The inner layer consists of the most critical elements of the system calculated at high-level of theory and the rest of the system comprises the outer layer calculated at a lower level of theory, which yield a consistent energy expression with similar accuracy to a high-level calculation in the full system [19]. In the terminology of Morokuma [20], the full system is called "real" treated with a low level of theory, and the inner layer is termed "model" treated with both the low and the high levels of theory. The total ONIOM energy (E^{ONIOM}) is expressed as Eq. (3) [22]:

$$E^{\text{ONIOM}} = E(\text{high, model}) + E(\text{low, real}) - E(\text{low, model}) \quad (3)$$

where $E(\text{high, model})$, $E(\text{low, real})$ and $E(\text{low, model})$ represent the energy of the inner layer ((R/S)-MMA) at the high level of theory, the entire system at the low level of theory (the complexes), and the energy of the model system (PM-β-CD) at the low level of theory, respectively.

In NBO analysis, the hyperconjugative interaction can be quantitatively described using the stabilization energy ($E^{(2)}$) which represents the estimate of the off-diagonal NBO Fock matrix elements [30]. This stabilization energy can be deduced from the second order perturbation approach [31]:

$$E^{(2)} = -n_\sigma \frac{\langle \sigma|F|\sigma \rangle}{\varepsilon_{\sigma*} - \varepsilon_\sigma} = -n_\sigma \frac{F_{i,j}^2}{\Delta E} = -n_\sigma \frac{F_{i,j}^2}{E_j - E_i} \quad (3)$$

where $\langle \sigma|F|\sigma \rangle$ or $F_{i,j}^2$ is the Fock matrix element between the i and j NBO orbitals, $\varepsilon_{\sigma*}$ and ε_σ are the energies of σ and σ* NBO orbitals, and n_σ is the population of the donor σ orbital. E_i, E_j are diagonal elements (orbital energies) and $F_{i,j}$ is the off-diagonal NBO Fock matrix element.

Results and discussion

Binding energies and structures of the inclusion complexes

PM3 method was adopted to search for the lowest energy structures of the inclusion complexation of (R/S)-MMA into PM-β-CD cavity. The results demonstrated the binding energies of the inclusion complexation of (R/S)-MMA into PM-β-CD cavity were changed at different positions (Z) and orientations. The more negative the binding energy is, the stronger the interaction between (R/S)-MMA and PM-β-CD. From Table 1, it can be seen that the binding energies of (R)-MMA/PM-β-CD and (S)-MMA/PM-β-CD complexes are lowest as (R/S)-MMA molecules were included into PM-β-CD by the orientation of type 3, which are −79.14 and −85.71 kJ mol^{-1}, respectively. It was evidenced that the complexation process is energetically favorable.

Table 1 Energies in four kinds of stabilization inclusion complexes of (*R/S*)-MMA with PM-β-CD calculated using PM3 method and insert positions of (*R/S*)-MMA

Inclusion type	(*R*)-MMA/PM-β-CD				(*S*)-MMA/PM-β-CD			
	E (kJ mol^{-1})	BE (kJ mol^{-1})	DEF (kJ mol^{-1})	Z (Å)	E (kJ mol^{-1})	BE (kJ mol^{-1})	DEF (kJ mol^{-1})	Z (Å)
type 1	−6026.53	−67.76	0.80	2.91	−6024.78	−65.81	2.80	2.92
type 2	−6017.44	−58.67	0.95	3.86	−6017.45	−58.48	4.40	4.78
type 3	−6037.90	−79.14	3.17	3.96	−6044.68	−85.71	2.97	4.17
type 4	−6008.76	−50.00	4.81	1.56	−6014.48	−55.51	4.80	1.26

E is the heat of formation of (*R/S*)-MMA/PM-β-CD complexes, BE is the binding energy, DEF is the deformation energy of (*R/S*)-MMA, Z is the final distance between the chiral carbon atom and the XY plane in optimized (*R/S*)-MMA/PM-β-CD complexes

The heats of formation of the optimized (*R*)-MMA and (*S*)-MMA should be the same in theory, as there is a difference for the heats of formation calculated using PM3 method, which is 0.21 kJ mol^{-1}. The density functional method (DFT) at B3LYP/6-31G(d) level was adopted to further optimize (*R/S*)-MMA, which provided equal values (−1508735.73 kJ mol^{-1}) for the guests of both chiralities. It is indicated that the accuracy of PM3 calculation is lower than that of DFT. DFT, as more advanced method, is necessary to re-optimize the geometries. However, the DFT is prohibitively expensive for optimization and frequency calculations of the large systems like cyclodextrin complexes. Therefore, the two-hybrid ONIOM method (B3LYP/6-31G(d):PM3) was adopted to fully optimize the geometries of the most stable (*R/S*)-MMA/PM-β-CD complexes obtained by PM3 method, either reducing the computational cost or increasing the accuracy of the resulting structure. The system was divided into two parts: the most important part consisting of (*R/S*)-MMA molecules for the inner layer; and the minor part consisting of the remaining part PM-β-CD for the outer layer. The total optimized energies (E^{ONIOM}) of the (*R*)-MMA/PM-β-CD and (*S*)-MMA/PM-β-CD complexes calculated by ONIOM method are −1514334.41 kJ mol^{-1} and −1514342.01 kJ mol^{-1}, respectively. The difference of the two binding energy on ONIOM method is slightly larger than that on PM3 method, as the energy differences ($\Delta(BE)$) for ONIOM and PM3 methods are 7.60 kJ mol^{-1} and 6.78 kJ mol^{-1}, respectively. These results follow the same trend as the PM3 optimizations, that the energy of the most stable (*S*)-MMA/PM-β-CD complex is more negative than that of the most stable (*R*)-MMA/PM-β-CD complex. It can be seen that the (*S*)-MMA/PM-β-CD complex is more stable than the (*R*)-MMA/PM-β-CD complex, that is to say, the interaction between (*S*)-MMA and PM-β-CD is stronger than that between (*R*)-MMA and PM-β-CD. It is consistent with the result of our previous report [9]. We had successfully separated the enantiomers of (*R/S*)-MMA on PM-β-CD CSP by gas chromatography and found that the retention time of (*S*)-MMA is longer than (*R*)-MMA. The longer the retention time is, the stronger the interaction between (*S*)-MMA and PM-β-CD CSP.

The optimized geometries of the inclusion complexes obtained by ONIOM method are presented in Fig. 2. Although (*R*)-MMA and (*S*)-MMA both lie on the wider edge of the PM-β-CD cavity, the binding geometries in the two complexes are fully different. The benzene ring of (*R*)-MMA locates horizontally approximately on the wider edge of the PM-β-CD cavity with the ester group includes into PM-β-CD cavity as the angle between the bond C7-C8 and the Z axis is 52.72°. On the contrary, the ester group of (*S*)-MMA locates horizontally approximately while the benzene ring is deeply included into the hydrophobic cavity as the angle between the bond C7-C5 and the Z axis is 38.93°. However, the carbonyls of (*R/S*)-MMA both point to the up (the wider edge of the PM-β-CD cavity). The distances between the chiral carbon of (*R*)-MMA and the O2 and O3 in the G1 glucose (as shown in Fig.1) are 5.07 and 5.48 Å, respectively, while the vertical distance between the chiral carbon and the XY plane is 4.57 Å. Relatively, the distances between the chiral carbon of (*S*)-MMA and the O2 and O3 in the G1 glucose (as shown in Fig.1) are 4.83 and 4.17 Å, respectively, while the vertical distance between the chiral carbon and the XY plane is 4.20 Å. We can note that, on the most stable structures, the chiral carbon in (*R/S*)-MMA molecule is close to C2、C3 (chiral carbon) in glucose units, and the interaction between (*S*)-MMA and PM-β-CD is stronger than (*R*)-MMA. It is indicated that the differences between (*R/S*)-MMA/PM-β-CD inclusion geometries are important for the chiral recognition (*R/S*)-MMA on PM-β-CD and the chiral recognition mechanism was closely related to the chiral environment provided by C2 and C3 in the glucose unit and the degree of (*R/S*)-MMA with PM-β-CD inclusion.

The linkage torsion angles ϕ(C4-O4-C1′-O5′) and ψ(C5-C4-O4-C1′), often mentioned as the primary variables of oligosaccharide structure, together with the O1···O4 distance ($d_{O1···O4}$) and the pseudo torsion angle ω(O1–C1···C4–O4), as simple descriptors of the various monomeric shapes, are listed in Table 2. It can be seen that the above four configuration parameters of PM-β-CD are

Fig. 2 Energy-minimized structures obtained by ONIOM calculations for the (R/S)-MMA/PM-β-CD complexes. (**a**) (R)-MMA/PM-β-CD seen from the side of the PM-β-CD wall; (**b**) (S)-MMA/PM-β-CD seen from the side of the PM-β-CD wall; (**c**) (R)-MMA/PM-β-CD seen from the wider edge of the PM-β-CD cavity; (**d**) (S)-MMA/PM-β-CD seen from the wider edge of the PM-β-CD cavity. The possible intermolecular hydrogen bonds are shown as a dotted line

(a) (R)-MMA/ PM-β-CD

(b) (S)-MMA/ PM-β-CD

(c) (R)-MMA/ PM-β-CD

(d) (S)-MMA/ PM-β-CD
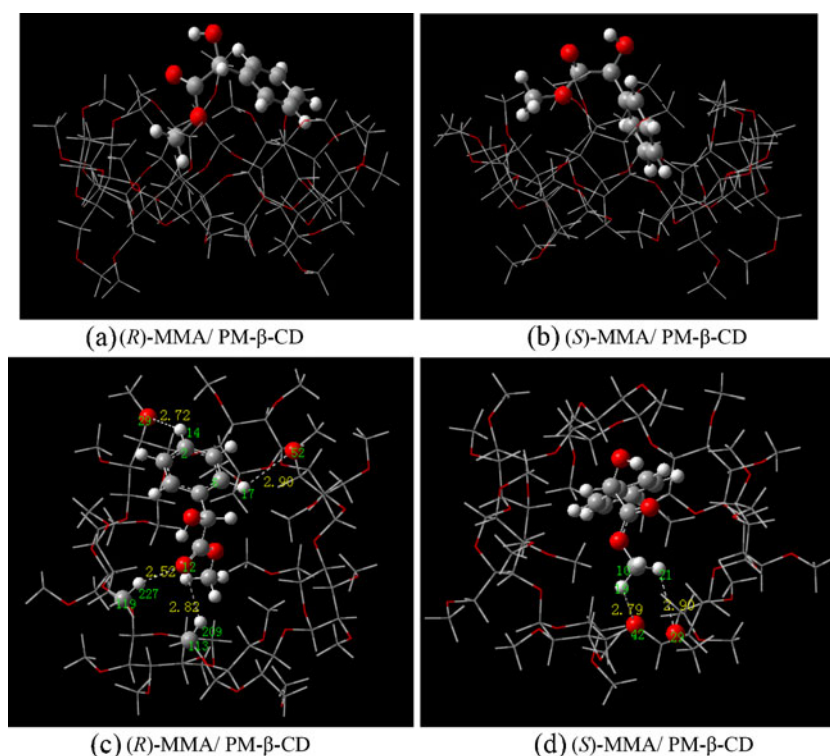
changed obviously after interacting with (R/S)-MMA. And the configurations of PM-β-CD in the two stable inclusion complexes are different. Similarly, the conformations of (R/S)-MMA molecules are changed due to complex formation, as shown in Table 3, indicating that the conformational flexibility of the (R/S)-MMA molecules mainly reflects on the bond angle changes. It suggests that conformational flexibility is a feature of (R/S)-MMA and PM-β-CD molecules. In a previous study of Rekharschy and Inoue [3], they stated that flexibility enhanced the complexation entropy since "more possible conformers can fit properly into the cavity" on their report. In other words, the flexibility of host and guest can favor the host–guest interaction because the host and guest can modify their conformations to ensure a better inclusion. On the other hand, the *DEF*s of the (R)-MMA and (S)-MMA molecules in the most stable complexes calculated by ONIOM method are 2.14 kJ mol^{-1} and 3.36 kJ mol^{-1}, respectively, which demonstrate that the (S)-MMA molecule requires slightly more energy than (R)-MMA in order to adapt its structure to bind within the PM-β-CD cavity. This can be supported by the fact that flexibility of the guest and host molecules plays an important role in increasing the stability of the whole system upon complexation.

Thermodynamics of binding process

The statistical thermodynamics calculations were performed using Harmonic frequency analysis in PM3

method for the most stable structures, characterizing them as true minima on the potential energy surface. The frequencies analyses were then used for the evaluation of the thermodynamic parameters, such as enthalpy changes ($\Delta H°$), entropy contribution ($\Delta S°$) and Gibbs free energy ($\Delta G°$), for the statistical thermodynamic parameters in binding process of (R/S)-MMA with PM-β-CD at 298.15 K at 1 atm were summarized in Table 4. It can be observed that the inclusion complexation of (R/S)-MMA with PM-β-CD are exothermic judged from the negative enthalpy changes. The enthalpy change for the (S)-MMA/PM-β-CD complex is more negative than that for the (R)-MMA/PM-β-CD complex. It means that both the inclusion processes are enthalpically favorable in nature due to the negative enthalpy changes [32]. The $\Delta H°$ and $\Delta S°$ in the inclusion process of (R/S)-MMA with PM-β-CD to form the most stable complexes are negative, suggesting that the formation of the inclusion complex is an enthalpy-driven process in vacuum and the conclusion agrees well with the experimental results obtained by gas chromatography [9]. The positive $\Delta G°$ value can be attributed to the underestimation of the calculated $\Delta S°$. It is mainly caused by the reason that the semiempirical calculations neglect the effect, which there is a gain in entropy due to the assimilation of the solvation molecules by the medium after inclusion takes place [33]. So, the Gibbs free energy obtained from semiempirical calculations has no absolute meaning and should be considered only in a comparative way.

Table 2 The primary variables of PM-β-CD conformation after interaction with (R/S)-MMA

	ϕ (°)	ψ (°)		$d_{O1\cdots O4}$ (Å)	ω (°)	Conformation
G1···G2	138.25	−85.58	G1	4.25	0.91	free
	121.71	−78.32		4.24	4.10	(R)-complex
	125.82	−76.03		4.20	1.28	(S)-complex
G2···G3	93.72	−118.07	G2	4.27	5.91	free
	93.68	−118.51		4.36	3.80	(R)-complex
	93.19	−121.77		4.33	6.55	(S)-complex
G3···G4	88.42	−141.31	G3	4.41	1.72	free
	85.63	−146.42		4.38	-2.19	(R)-complex
	84.69	−145.43		4.42	-2.01	(S)-complex
G4···G5	136.24	−76.72	G4	4.25	6.31	free
	133.19	−73.55		4.30	9.10	(R)-complex
	131.53	−73.68		4.29	8.35	(S)-complex
G5···G6	76.81	−150.64	G5	4.25	−3.83	free
	76.80	−149.22		4.19	−3.36	(R)-complex
	75.40	−146.66		4.06	−5.33	(S)-complex
G6···G7	130.60	−73.26	G6	4.13	−0.28	free
	132.20	−70.21		4.08	−0.32	(R)-complex
	134.81	−77.89		4.24	0.85	(S)-complex
G7···G1	75.17	−151.78	G7	4.18	−4.61	free
	76.64	−152.37		4.32	−4.38	(R)-complex
	79.17	−147.17		4.28	−1.75	(S)-complex

G i (i=1-7) refers to the number of the D-glucosyl units of PM-β-CD; ϕ, ψ and ω refer to the torsion angle of C4-O4-C1′-O5′, C5-C4-O4-C1′ and O1···C1···C4–O4, respectively; the numbers of the above atoms are shown in Fig. 1; free, R-complex and S-complex refer to the free PM-β-CD, (R)-MMA/PM-β-CD complex and (S)-MMA/PM-β-CD

NBO analysis

In NBO analysis, the electronic wavefunctions are interpreted in terms of a set of occupied Lewis and a set of

Table 3 The primary bond distances and angles of free and complexed (R/S)-MMA into PM-β-CD

	S-MMA		R-MMA	
	Free	Sp.(ONIOM)	Free	Sp.(ONIOM)
Bond distances (nm)				
C5-C7	0.1530	0.1525	0.1530	0.1523
C7-C8	0.1535	0.1531	0.1535	0.1535
C8-O9	0.1336	0.1331	0.1336	0.1334
O9-C10	0.1442	0.1438	0.1442	0.1441
C7-O11	0.1408	0.1410	0.1408	0.1409
C8-O12	0.1216	0.1219	0.1216	0.1218
O11-H22	0.0977	0.0977	0.0977	0.0976
Angles (°)				
C5-C7-C8	110.30	108.97	110.30	111.00
C7-C8-O9	113.13	112.63	113.13	113.32
C8-O9-C10	115.76	116.47	115.76	116.07
O11-C7-C5	112.73	112.18	112.73	112.13
O11-C7-O8	108.18	108.45	108.18	108.29
C7-O11-H22	108.14	105.47	107.41	105.76

Free represents the free guest, and Sp. represents the guest from the complex. The atomic numbering scheme for atoms were depicted in Fig. 1

unoccupied non-Lewis localized orbitals. The delocalization of electron density between occupied Lewis-type NBO orbital and formally unoccupied non-Lewis NBO orbital corresponds to a stabilizing electron donor-acceptor interaction. So, the information about interaction in both filled and virtual orbital obtained by the NBO analysis could help the analysis of intra- and intermolecular interactions. The stabilization energy ($E^{(2)}$) is related to the delocalization trend of electrons from the bonding or nonbonding orbitals to the antibonding orbitals, which is estimated by second order perturbation theory [34]. NBO analysis has been proved to be an effective tool for investigating hyperconjugative interaction and electron density transfer (EDT) from LP(Y) of the "Lewis base" Y into BD* (X–H) of the "Lewis acid" X–H in X–H···Y hydrogen bonding systems [35, 36]. Therefore, NBO analysis was performed on (R/S)-MMA/PM-β-CD complexes in order to elucidate the intramolecular and intermolecular hydrogen bonding, intermolecular charge transfer and dipole-dipole interaction between the host and guest. For the most stable (R/S)-MMA/PM-β-CD complexes optimized by ONIOM method, the stabilization energies ($E^{(2)}$) associated with the primary hyperconjugative interactions between a lone pair [LP(O)] of an atom O, and an antibonding orbital [BD*(C−H)] were list in Table 5, which quantify the extent of intramolecular and intermolecular hydrogen bonding for (R/S)-MMA/PM-β-CD complexes. It is suggested, in general, that the $E^{(2)}$ value is larger than 8.37 kJ mol^{-1} for strong

Table 4 The thermodynamic parameters upon the most stable inclusion complexation of PM-β-CD with (R/S)-MMA by PM3 method at 298.15 K at 1 atm

	$\Delta H°$(kJ/mol)	$\Delta S°$(J mol^{-1}K^{-1})	$\Delta G°$(kJ/mol)
(R)-MMA/ PM-β-CD	−67.19	−315.84	26.93
(S)-MMA/ PM-β-CD	−71.51	−274.35	10.25

hydrogen bonding interaction and from 8.37 kJ mol^{-1} to 2.09 kJ mol^{-1} for weak hydrogen bonding interaction [37]. As shown in Table 5, the stabilization energy $E^{(2)}$ of intermolecular C–H⋯O hydrogen bonding in both inclusion complexes are smaller than 8.37 kJ mol^{-1}, indicating that there are two weak hydrogen bondings and two very weak hydrogen bondings in (R)-MMA/PM-β-CD inclusion complex, as there are merely two very weak hydrogen bondings in (S)-MMA/PM-β-CD inclusion complex. The O⋯H distances and C–H⋯O angles of these hydrogen bondings range from 2.72 to 2.90 Å and 120.6 to 141.5 °, respectively, which just falls in the reported data (less than 3.0 Å and greater than 90° [38]). The detailed intermolecular hydrogen bonding interactions for the two inclusion complexes are shown in Fig. 2 as dotted lines. It is indicated that the binding geometries in the two complexes are both affected by several intermolecular C–H⋯O hydrogen bondings formed by the orbital overlap between LP(O) and BD*(C-H).

In addition, it is also observed in Table 5 that there is no delocalization of electron density between O–H moiety of (R/S)-MMA and O atoms of PM-β-CD molecule as there is delocalization of electron density between O–H moiety of (R/S)-MMA and C=O moiety of (R/S)-MMA. The stabilization energies $E^{(2)}$ of the intramolecular O–H⋯O hydrogen bondings in (R/S)-MMA molecule are larger than 19.99 kJ mol^{-1} and increase significantly with the formation of (R/S)-MMA/PM-β-CD complexes. It was indicated that the guest's hydroxyl group only formed strong intramolecular hydrogen bonding with the carbonyl group, and the hydrogen bonding interaction becomes stronger with the formation of (R/S)-MMA/PM-β-CD complexes.

Additionally, it is interesting to find out that although the hydrogen bonding interaction in (R)-MMA/PM-β-CD complex is stronger than that in (S)-MMA/PM-β-CD complex, the geometry of (R)-MMA/PM-β-CD complex is not more stable than that of (S)-MMA/PM-β-CD complex. To further verify their stability, the charge transfer between (R/S)-MMA and PM-β-CD for the most stable (R/S)-MMA/PM-β-CD complexes is investigated. The Mulliken charge of heavy atoms of (R/S)-MMA, charge transfer and dipole moment of (R/S)-MMA/PM-β-CD complexes calculated by NBO analysis at B3LYP/6-31G(d) level are summarized in Table 6. It can be seen from Table 6 that the charge transfers of (R)-MMA and (S)-MMA in complexes are 0.0166 e and −0.0239 e, respectively, indicating that (R)-MMA and (S)-MMA act as electron donor and electron acceptor, respectively. However, in term of the rough estimation 0.0001 electron of charge transfer corresponds to 4.187 kJ mol^{-1} of the stabilization energy [39], the stabilization energy of (S)-MMA/PM-β-CD produced by charge transfer is greater than that of (R)-MMA/PM-β-CD. From Fig. 2, it can also be seen that the benzene ring of (S)-MMA is deeply included in the hydrophobic cavity of PM-β-CD, so there is a stronger hydrophobic interaction between (S)-MMA and PM-β-CD. It was demonstrated that the hydrogen bonding interaction, dipole-dipole interaction, charge transfer and hydrophobic interaction all play important roles in the process of formation of (R/S)-MMA/PM-β-CD complexes.

In an earlier paper, Lipkowitz and Stoehr [14] examined the binding of (R/S)-MMA to β-Cyclodextrin in great detail. They used chromatography, NMR spectroscopy, and molecular dynamics simulations to study the guest-host complexation between MMA enantiomers with β-Cyclodextrin, and some conclusions were drawn. First, the more tightly bound enantiomer is the S-isomer which is the same as our result. Second, short range dispersion forces rather than long range Coulombic forces are responsible for both complexation and for enantiodiscrimination. Third, the intermolecular hydrogen bonds are not discriminating, whereas we find that there are weak hydrogen bonding interaction, dipole-dipole interaction, charge-transfer and hydrophobic interaction between (R/S)-MMA enantiomers and PM-β-CD by using NBO analysis. So, the main driven forces responsible for chiral recognition are weak hydrogen bonding interaction, dipole-dipole interaction, charge transfer and hydrophobic force.

Conclusions

The complexes of the MMA enantiomer inserted within PM-β-CD were fully characterized theoretically. The results show clearly that the third orientation (type 3) of (R/S)-MMA inside PM-β-CD is more favorable than the other three orientations. And the interaction between (S)-MMA and PM-β-CD is stronger in agreement with the result of gas chromatograph observation. Higher levels of theory resulted in a more reasonable structure for the complex, although the difference in energy between the optimized structures of the most stable (R)-MMA/PM-β-CD and (S)-MMA/PM-β-CD complexes are not large. The optimized geometries of (S)-MMA/PM-β-CD and (R)-

Table 5 The partial electron donor and acceptor orbitals and corresponding $E^{(2)}$, distances and angles calculated using NBO analysis for (R/S)-MMA/PM-β-CD complexes optimized by ONIOM method

Electron donor (i)	Electron acceptor (j)	$E^{(2)}$ [a] (kJ/mol)	$E(j)-E(i)$ [b] (a.u.)	$F(i,j)$ [c] (a.u.)	$d_{O···H}$ (Å)	Angel [d] (°)
Free (R/S)-MMA						
LP(1) O12	BD*(1) O11-H22	4.73	1.14	0.032	2.02	120.16
LP(2) O12	BD*(1) O11-H22	19.99	0.72	0.054		
(R)-MMA/PM-β-CD complex						
Within (R)-MMA						
LP(1) O12	BD*(1) O11-H22	5.35	1.15	0.034	2.01	120.70
LP(2) O12	BD*(1) O11-H22	20.70	0.73	0.055		
from (R)-MMA to PM-β-CD						
LP(1) O12	BD*(1) C113 - H209	0.2927	1.18	0.008	2.82	118.49
LP(2) O12	BD*(1) C113 - H209	0.7946	0.76	0.011		
LP(1) O12	BD*(1) C119 - H227	3.3456	1.17	0.027	2.52	158.50
LP(2) O12	BD*(1) C119 - H227	0.9619	0.75	0.012		
from PM-β-CD to (R)-MMA						
LP(1) O29	BD*(1) C2 - H14	2.0910	1.07	0.021	2.72	141.53
LP(2) O29	BD*(1) C2 - H14	0.5018	0.80	0.009		
LP(1) O52	BD*(1) C6 - H17	0.2091	1.05	0.007	2.90	124.41
(S)-MMA/PM-β-CD complex						
Within (R)-MMA						
LP(1) O12	BD*(1) O11-H22	4.98	1.14	0.033	2.01	120.79
LP(2) O12	BD*(1) O11-H22	21.29	0.72	0.056		
from PM-β-CD to (S)-MMA						
LP(2) O29	BD*(1) C10 - H21	0.2091	0.76	0.006	2.90	105.24
LP(1) O42	BD*(1) C10 - H19	0.7109	1.04	0.012	2.79	120.64
LP(2) O42	BD*(1) C10 - H19	0.4182	0.79	0.008		

[a] $E^{(2)}$ means energy of hyperconjugative interactions; cf. Eq. (3); [b] Energy difference between donor (i) and acceptor (j) NBO orbitals; [c] $F(i,j)$ is the Fock matrix element between i and j NBO orbitals; [d] The angel of O···H-C. LP denotes valence lone pair; BD denotes bonding orbital; BD* denotes σ* antibonding orbital. For LP, (1) and (2) denote the first and second lone pair electron, respectively. For DB and BD*, (1) denotes σ orbital, (2) denotes π orbital

Table 6 Mulliken charges (**e**) of the heavy atoms, charge transfer and dipole moment of (R/S)-MMA in complexes calculated by NBO analysis at B3LYP/6-31G(d) level

Atoms	free (R/S)-MMA	In (R/S)-MMA/PM-β-CD complexes	
		(R)-MMA	(S)-MMA
C1	-0.0046	0.0068	0.0019
C2	0.0024	0.0202	-0.0002
C3	-0.0010	-0.0092	-0.0062
C4	-0.0217	-0.0289	-0.0368
C6	0.1305	0.1534	0.1657
C6	-0.0354	-0.0274	-0.0384
C7	0.1677	0.1446	0.1339
C8	0.6044	0.6189	0.6186
O9	-0.4392	-0.4396	-0.4280
C10	0.3062	0.2983	0.2933
O11	-0.2213	-0.2233	-0.2234
O12	-0.4879	-0.4972	-0.5044
Charge transfer	0.0000	0.0166	-0.0239
Dipole moment (Debye)	2.9317	3.0564	3.1217

MMA/PM-β-CD complex are different. The benzene ring of (R)-MMA locates horizontally approximately on the wider edge of the PM-β-CD cavity, but the benzene ring of (S)-MMA is deeply included into the hydrophobic cavity. The results obtained using NBO calculations show that the main driving forces in the process complex are weak hydrogen bonding interaction, dipole-dipole interaction, charge-transfer function and hydrophobic interaction. Moreover, in (R/S)-MMA/PM-β-CD inclusion complex, the chiral carbon of (R/S)-MMA are both close to C2 and C3 in the glucose units, thus the chiral selector capacity is mainly due to the chiral environment provided by C2 and C3 in the glucose unit and the tightness of combining between (R/S)-MMA and PM-β-CD.

References

1. Saenger W, Jacob J, Gessler K, Steiner T, Hoffmann D, Sanbe H, Koizumi K, Smith SM, Takaha T (1998) Structures of the common cyclodextrins and their larger analogues beyond the doughnut. Chem Rev 98:1755–1786
2. French AD, Johnson GP (2007) Linkage and pyranosyl ring twisting in cyclodextrins. Carbohydr Res 342:1223–1237
3. Rekharsky MV, Inoue Y (1998) Complexation thermodynamics of cyclodextrins. Chem Rev 98:1875–1917
4. Lipkowitz KB (1998) Applications of computational chemistry to the study of cyclodextrins. Chem Rev 98:1829–1873
5. Xing SK, Zhang C, Ai HQ, Zhao Q, Zhang Q, Sun DZ (2009) Theoretical study of the interactions of β-cyclodextrin with 2′-hydroxyl-5′-methoxyacetophone and two of its isomers. J Mol Liq 146:15–22
6. Beier T, Höltje HD (1998) Modified cyclodextrins as chiral selectors: molecular modelling investigations on the enantioselective binding properties of heptakis(2,3-di-O-methyl-6-O-tert-butyldimethylsilyl)-β-cyclodextrin. J Chromatogr B 708:1–20
7. Shahgaldian P, Pieles U (2006) Cyclodextrin derivatives as chiral supramolecular receptors for enantioselective sensing. Sensors 6:593–615
8. Juvancz Z, Kendrovics RB, Iványi R, Szente L (2008) The role of cyclodextrins in chiral capillary electrophoresis. Electrophoresis 29:1701–1712
9. Shi JH, Cheng XW, Yan W (2009) Enantiomeric separation of methylmandelate by gas chromatography and discussion the chiral recognition mechanism. Chin J Pharm Anal 29:1681–1684
10. Grandeury A, Petit S, Gouhier G, Agasse V, Coquerel G (2003) Enantioseparation of 1-(p-bromophenyl)ethanol by crystallization of host–guest complexes with permethylated β-cyclodextrin: crystal structures and mechanisms of chiral recognition. Tetrahedron: Asymmetry 14:2143–2152
11. Harata K (1998) Structural aspects of stereodifferentiation in the solid state. Chem Rev 98:1803–1827
12. Kano K, Kato Y, Kodera M (1996) Mechanism for chiral recognition of binaphthyl derivatives by cyclodextrins. J Chem Soc Perkin Trans 2:1211–1217
13. Ding HY, Chao JB, Zhang GM, Shuang SM, Pan JH (2003) Preparation and spectral investigation on inclusion complex of b-cyclodextrin with rutin. Spectrochim Acta A 59:3421–3429
14. Lipkowitz KB, Stoehr CM (1996) Detailed experimental and theoretical analysis of chiral discrimination: enantioselective

15. Dos Santos HF, Duarte HA, Sinisterra RD, De Melo Mattos SV, De Oliveira LFC, De Almeida WB (2000) Quantum-mechanical study of the interaction of α-cyclodextrin with methyl mercury chloride. Chem Phys Lett 319:569–575
16. Liu L, Guo QX (2004) Use of quantum chemical methods to study cyclodextrin chemistry. J Incl Phenom Macro 50:95–103
17. Yan CL, Xiu ZL, Li XH, Hao C (2007) Molecular modeling study of β-cyclodextrin complexes with (+)-catechin and (−)-epicatechin. J Mol Graph Model 26:420–428
18. Maseras F, Morokuma K (1995) Imomm - a new integrated ab initio plus molecular mechanics geometry optimization scheme of equilibrium structures and transition-states. J Comput Chem 16:1170–1179
19. Svensson M, Humbel S, Froese RDT, Matsubara T, Sieber S, Morokuma K (1996) ONIOM: a multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. a test for diels-alder reactions and Pt(P(t-Bu)3)2+H2 oxidative addition. J Phys Chem 100:19357–19363
20. Dapprich S, Komaromi I, Byun KS, Morokuma K (1999) A new ONIOM implementation in Gaussian98. I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. J Mol Struct THEOCHEM 461(462):1–21
21. Holt JS (2010) Structural characterization of the Brooker's merocyanine/β-cyclodextrin complex using NMR spectroscopy and molecular modeling. J Mol Struct 965:31–38
22. Yan CL, Xiu ZL, Li XH, Teng H, Hao C (2007) Theoretical study for quercetin/β-cyclodextrin complexes: quantum chemical calculations based on the PM3 and ONIOM2 method. J Incl Phenom Macro 58:337–344
23. Tafazzoli M, Ghiasi M (2009) Structure and conformation of α-, β- and γ-cyclodextrin in solution: Theoretical approaches and experimental validation. Carbohyd Polym 78:10–15
24. Chen HY, Ji HB (2010) Alkaline hydrolysis of cinnamaldehyde to benzaldehyde in the presence of β-Cyclodextrin. AIChE J 56:466–476
25. Wang LF (2009) Halogenation effects of pheniramines on the complexation with β-cyclodextrin. J Pharm Biomed Anal 50:392–396
26. Nie MY, Zhou LM, Wang QH, Zhu DQ (2000) Gas chromatographic enantiomer separation of mandelates and its analogs on permethylated cyclodextrin chiral stationary phases. Chin J Anal Chem 28:1366–1370
27. Foresman JB, Frisch A (1996) Exploring chemistry with electronic structure methods, 2nd edn. Pittsburgh, PA, USA
28. Shi JH, Xiao KK, Lü YY (2009) Host-guest interactions of β-cyclodextrin with enantiomers of ethyl α-chloropropionates. Acta Phys Chim Sin 25:1273–1278
29. Yahia OA, Khatmi DE (2009) Theoretical study of the inclusion processes of Venlafaxine with β-cyclodextrin. J Mol Struct THEOCHEM 912:38–43
30. Li XH, Yin GX, Zhang XZ (2010) Natural bond orbital (NBO) population analysis of some benzyl nitrites. J Mol Struct (THEOCHEM) 957:61–65
31. Chocholoušová J, Špirko V, Hobza P (2004) First local minimum of the formic acid dimer exhibits simultaneously red-shifted O–H…O and improper blue-shifted C–H…O hydrogen bonds. Phys Chem Chem Phys 6:37–41
32. Yan L, Li XH, Xiu ZL, Hao C (2006) A quantum-mechanical study on the complexation of β-cyclodextrin with quercetin. J Mol Struct THEOCHEM 764:95–100
33. Saenger W (1980) Cyclodextrin inclusion compounds in research and industry. Angew Chem Int Edn Engl 19:344–362
34. Reed AE, Curtiss LA, Weinhold F (1988) Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. Chem Rev 88:899–926

35. Mukherjee V, Singh NP, Yadav RA (2011) Optimized geometry and vibrational spectra and NBO analysis of solid state 2,4,6-tri-fluorobenzoic acid hydrogen bonded dimer. J Mol Struct 988:24–34

36. Madi F, Khatmi D, Dhaoui N, Bouzitouna A, Abdaoui M, Boucekkine A (2009) Molecular model of CENS piperidine β-CD inclusion complex: DFT study. C R Chimie 12:1305–1312

37. Uccello-Barretta G, Balzano F, Sicoli G, Fríglola C, Aldana I, Monge A, Paolino D, Guccione S (2004) Combining NMR and molecular modelling in a drug delivery context: investigation of the multi-mode inclusion of a new NPY-5 antagonist bromoben-zenesulfonamide into β-cyclodextrin. Bioorgan Med Chem 12:447–458

38. Van den Worm E, Beukelman CJ, van den Berg AJJ, Kroes BH, Labadie RP, van Dijk H (2001) Effects of methoxylation of apocynin and analogs on the inhibition of reactive oxygen species production by stimulated human neutrophils. Eur J Pharmacol 433:225–230

39. Chocholousova J, Spirko V, Hobza P (2004) First local minimum of the formic acid dimer exhibits simultaneously red-shifted O-H center dot center dot center dot O and improper blue-shifted C-H center dot center dot center dot O hydrogen bonds. Phys Chem Chem Phys 4:37–41

ORIGINAL PAPER

XRD and DFT-modelized structures of a pteridine-2,4 (1H,3H)-dithione/N,N′-dimethylformamide H-bonded cluster (2:2). MO study of the coordination ability

Nuria A. Illán-Cabeza · Tomás Peña-Ruiz · Miguel N. Moreno-Carretero

Abstract The title compound, $C_6H_4N_4S_2 \cdot C_3H_7NO$, crystallizes in the monoclinic space group C 2/c with $a=26.673(5)$, $b=5.397(1)$, $c=16.522(3)$ Å, $\beta=95.49(3)°$, $Z=8$, $R=0.0461$ for 1891 reflections with $I>2\sigma(I)$ and 174 parameters (4 restraints). Single pteridine-2,4(1H,3H)-dithione and dimethylformamide molecules are packed *via* N-H···O and N-H···N hydrogen bonds into centrosymmetric clusters containing two molecules of each class; these are roughly planar and placed into two different sets of planes -both containing the [−1,0,2] direction- mutually angled by 77.8°. Despite the distance between two neighbor planes in each set is *ca.* 3.4 Å, the analysis of π,π-stacking interactions shows too large slippage distance between aromatic rings from contiguous planes. Additional σ-π interactions between S2, S4 and O1S atoms and pyrazine or pyrimidine rings from adjacent molecules are present. The structure for the cluster [DTLM-DMF]$_2$ has been simulated by using the density functionals B1B95 (6-31 G(d) and 6-31+G(d) basis sets) and M06-2X (6-31 G(d) basis set). As a result, the M06-2X/6-31 G(d) approach provides the best agreement with the experimental XRD data. For a better evaluation of the intermolecular interactions, the superposition of two dimeric adducts [DTLM-DMF]$_2$ has been modelized. The binding capability of DTLM ligand was simulated on systems containing two metal-binding modes to palladium (N5-S4 and N1-S2) with different chelate size. The analysis of the frontier orbitals points out that the link with the metallic centers will take place through the sulfur atoms.

N. A. Illán-Cabeza · M. N. Moreno-Carretero (✉)
Departamento de Química Inorgánica y Orgánica,
Universidad de Jaén,
23071 Jaén, Spain
e-mail: mmoreno@ujaen.es

T. Peña-Ruiz
Departamento de Química Física y Analítica,
Universidad de Jaén,
23071 Jaén, Spain

Introduction

The pteridine systems are heterocyclic structures of coenzymatic constituents of oxidoreductase enzymes. They play essential roles in growth processes and the metabolism of one-carbon units [1] and are in clinical use as anticancer, antiviral, antibacterial, and diuretic drugs [2]. The investigation of several other enzymes involved in folate biosynthesis or in pteridine metabolism might be therapeutic targets so the variation of the substituents has provided new synthetic and highly functionalized compounds with interesting properties. Pteridine analogs have been evaluated as inhibitors for monoamine oxidase B and nitric oxide synthetase -which are involved in the pathology of neurodegenerative diseases [3]- and hepatitis C virus NS5B RNA-dependent RNA polymerase [4]; also, 4-amino-substituted pteridines have been tested as a new kind of nonnucleoside adenosine kinase inhibitors [5].

The sulfur-derivatives have been less studied. The 2,4-dithiolumazine (pteridine-2,4(1H,3H)-dithione, hereafter denoted as DTLM) was previously described and studied by means of pK$_a$ values and UV spectra [6].

Herewith, in order to complete previous papers already published [7–9] on the structure and metallation properties of lumazine derivatives, the crystal and molecular structure

Fig. 1 An ORTEP view of the H-bonded dimeric unit of the DTLM·DMF displaying the atom labeling scheme (ellipsoids at 50% probability). Hydrogen-bond details (D···A, Å; D-H···A,°) as follows: N(1)-H···N8 (1-x,1-y,1-z): 2.916(3), 179(5); N3-H···O1S (x,2-y,-1/2+z): 2.835(3), 177(3)
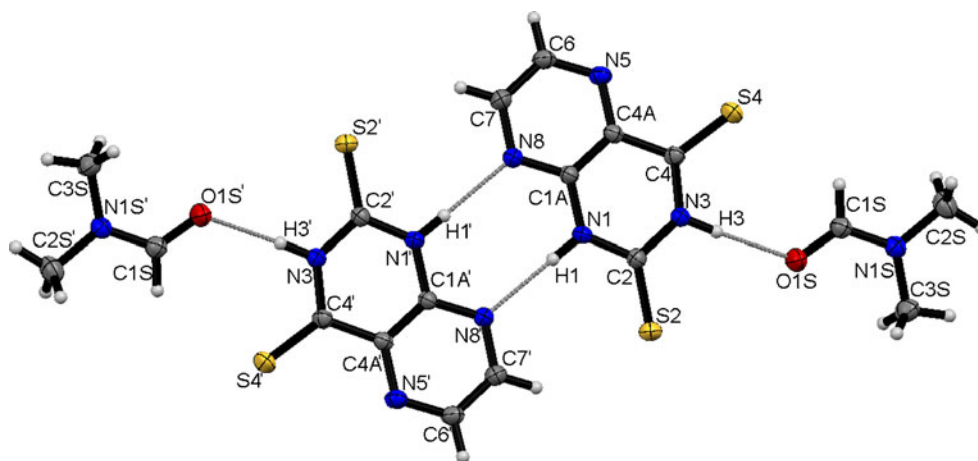
of the hydrogen-bonded cluster with pteridine-2,4(1*H*,3*H*)-dithione (DTLM) and dimethylformamide (DMF) (2:2), as well as a theoretical study on the intermolecular interactions are reported. Also, in order to elucidate the coordinations abilities of this pteridine derivative, a modelization of palladium compounds with different metal-ligand binding modes as well as an analysis of the frontier orbitals have been made.

Experimental

Synthesis

The pteridine compound (DTLM) was synthesized from the condensation of 5,6-diaminopyrimidine-2,4-thiol and glioxal following the Gabriel-Colman's method [10, 11]. The analytical data were: $C_6H_4N_4S_2$ (196.24 g·mol^{-1}), found 36.90% C, 2.91% H, 28.40% N, 33.06% S, calcd. 36.72% C, 3.08% H, 28.55% N, 32.67% S. Main IR data (cm^{-1}): 3038 ν(N-H) medium; 1576 ν(C=N) medium; 1481, 1251 ν(C=C)+ν(C-N) strong; 872, 824 ν(C=S) (thioamide IV) weak. ^{1}H-NMR data (δ, ppm): 8.74 d, 8.61 d (C-H); 13.63 s, 13.94 s (N-H). ^{13}C-NMR data (δ, ppm): 172.71 (C2), 187.67 (C4), 133.44 (C4a), 142.53 (C6), 149.20 (C7), 144.68 (C8a). Useful XRD red single-crystals of the DTLM·DMF adduct were isolated on recrystallizing the solid from a DMF:CH$_3$CN (1:1) solution.

Apparatus

Microanalyses of C, H, N and S were performed in a Termofinnigan Flash 1112 apparatus. IR spectra were obtained on a Bruker Tensor27 (KBr pellets, 4000–400 cm^{-1}) machine. ^{1}H and ^{13}C-NMR spectra were recorded on a Bruker DPX-400 spectrometer, using DMSO-d$_6$ as solvent and TMS as internal standard.

Crystallographic work

The measurements were performed on a Bruker-Nonius Kappa CCD diffractometer with graphite monochromated Mo-Kα radiation. Lorentz, polarization and multi-scan absorptions corrections were applied with *SADABS* [12] (max. and min. transmission, 0.9572 and 0.7935). The unit cell was determined from 28778 ramdom reflections (3.04<θ < 26.5º). The structure was solved by direct methods using *SHELXL97* program inside the *WINGX* package [13] employing full-matrix least-squares methods on F^2 [14]. Non-H atoms were located in successive difference Fourier maps and refined

Table 1 Bond lengths (Å) and angles (º)

|			
S2-C2	1.652(3)	N8-C7	1.327(4)
S4-C4	1.640(3)	C4A-C1A	1.394(4)
N1-C1A	1.367(3)	C4A-C4	1.459(4)
N1-C2	1.344(4)	C7-C6	1.390(4)
N5-C4A	1.333(3)	O1S-C1S	1.225(4)
N5-C6	1.322(4)	N1S-C1S	1.322(4)
N3-C2	1.373(4)	N1S-C2S	1.445(4)
N3-C4	1.365(4)	N1S-C3S	1.447(4)
N8-C1A	1.336(3)		
C1A-N1-C2	123.6(2)	S2-C2-N3	121.3(2)
C4A-N5-C6	116.7(2)	N1-C2-N3	115.9(2)
C2-N3-C4	126.6(2)	S4-C4-N3	121.1(2)
C1A-N8-C7	115.7(2)	S4-C4-C4A	123.9(2)
N5-C4A-C1A	121.0(2)	N3-C4-C4A	115.0(2)
N5-C4A-C4	120.5(2)	N8-C7-C6	122.0(3)
C1A-C4A-C4	118.4(2)	N5-C6-C7	122.0(3)
N1-C1A-N8	117.6(2)	O1S-C1S-N1S	125.3(3)
N1-C1A-C4A	120.1(2)	C1S-N1S-C2S	121.4(3)
N8-C1A-C4A	122.3(2)	C2S-N1S-C3S	117.2(3)
S2-C2-N1	122.8(2)	C3S-N1S-C1S	121.3(3)
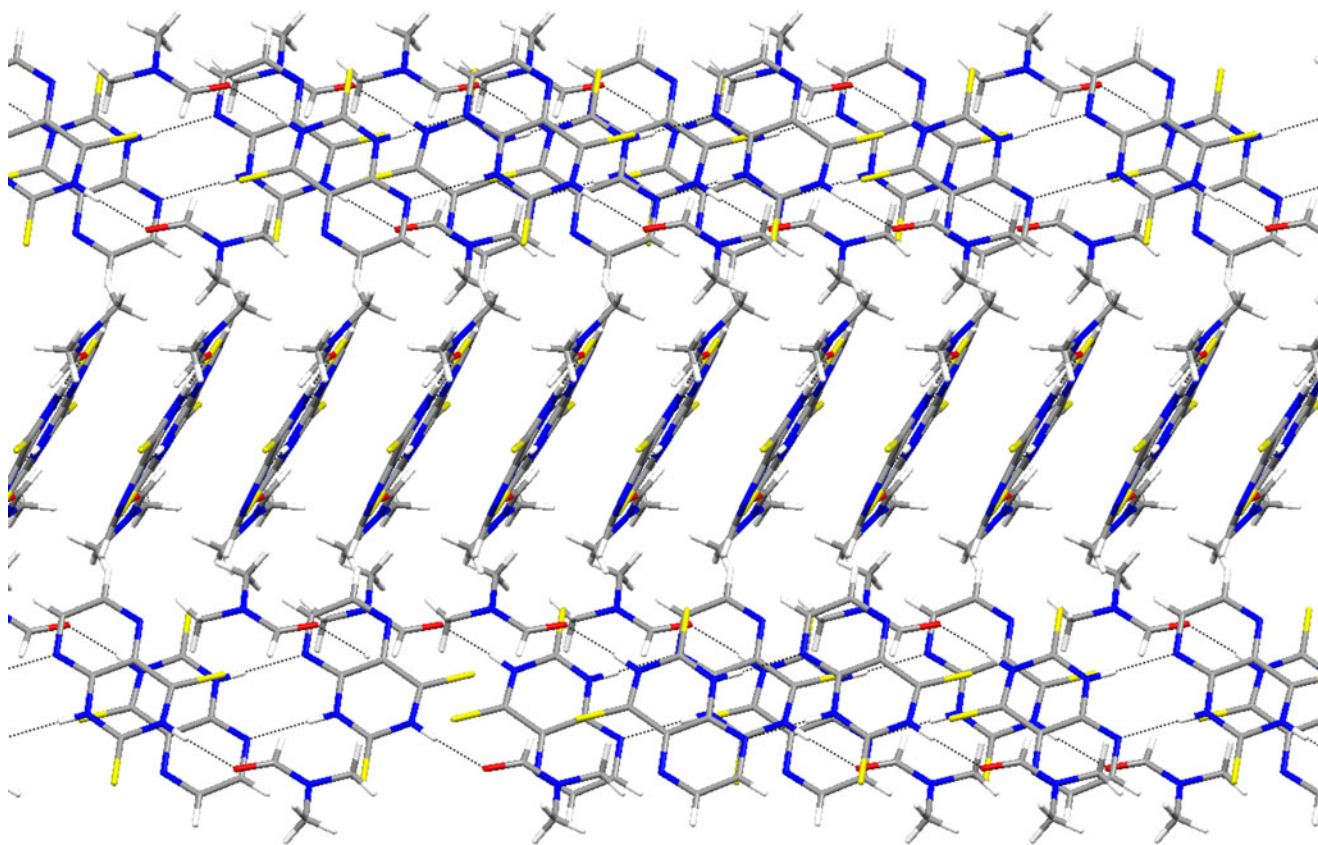
Fig. 2 A view of the crystal structure from the [150] direction, showing the two different orientation of the molecules

anisotropically. The H atoms attached to N1, N3, C6 and C7 were located in a difference Fourier map and were refined with a distance restraint of N-H=0.88(2) Å and C-H=0.95(2) Å, respectively. Methyl hydrogens were fixed geometrically and treated as riding with $U_{iso}=$ 1.5U_{eq}(C). Additional geometrical calculations were performed using *PLATON* [15] and drawings were carried out with *MERCURY* [16].

Crystal and refinement data: CCDC number 756140: Formula moiety $C_6H_4N_4S_2 \cdot C_3H_7NO$; M=269.35; Mono-

Fig. 3 Partial view of the H-bonds (black) and the σ-π (labeled Å, red) interactions in the crystal structure of the title compound. Centroids of pyrimidine and pyrazine rings are symbolized as light-gray balls
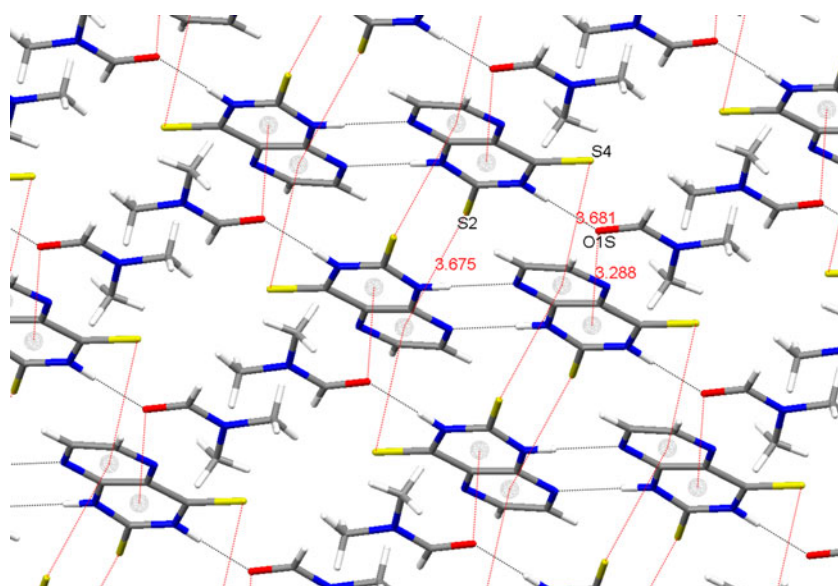
Table 2 RMSs for DTLM geometrical parameters

Calculation method	X-Y (Å)	X-Y-Z (°)
B1B95 6-31 G(d)	0.008	1.4
B1B95 6-31+G(d)	0.009	1.4
B1B95 cc-pVDZ	0.008	1.3
B1B95 aug-cc-pVTZ	0.008	1.3
B3LYP 6-31 G(d)	0.014	1.5
B3LYP 6-31+G(d)	0.015	1.4
B3LYP cc-pVDZ	0.015	1.3
B3LYP aug-cc-pVTZ	0.010	1.3
MP2(full) 6-31 G(d)	0.014	1.7
MP2(full) 6-31+G(d)	0.015	1.7
MP2 cc-pVDZ	0.019	1.7
MP2 aug-cc-pVTZ	0.011	1.4

clinic, C2/c; unit cell $a=26.673(5)$, $b=5.397(1)$, $c=16.522$ (3) Å; $\beta=95.49(3)°$, $V=2367.3(8)$ Å^3; $T=293(2)$ K; Bruker-Nonius KappaCCD apparatus, graphite-monochromated Mo-Kα radiation ($\lambda=0.71073$ Å); $Z=8$; $D_x=1.511$ Mg·m^{-3}; $F(000)=1120$; red prismatic crystal (0.55 × 0.27 × 0.10 mm^3); $\mu=0.441$ mm^{-1}; $\theta=3.04$-$26.5°$, $-33<h<33$, $-6<k<6$, $-20<l<20$; 28778 measured reflections, 2451 independent, 1891 with $I>2\sigma(I)$ used in the refinement ($R_{int}=0.0380$). Weighting scheme $w^{-1}=\sigma^2(F_o^2)+(0.0688P)^2+6.0096P$, where $P=1/3$ $(F_o^2+2F_c^2)$. Final R_1 and wR^2 [$I>2\sigma(I)$], 0.0461 and 0.1244; final R_1 and wR^2 [all data], 0.0657 and 0.1465; data-to-parameter ratio=14.1 (4 restraints); gof=1.042. Max. and min. $\Delta\rho$ in the final difference Fourier map were 0.363 and -0.368 e·Å^{-3}.

Computational details

Gaussian03 package [17] has been used to perform all the calculations in this research.

Becke's one and three parameters hybrid exchange functionals [18, 19] (B1 and B3) along with Becke's τ-dependent gradient-corrected (B95) [19] and Lee, Yang, and Parr's (LYP) [20] correlation functionals, respectively, have been used to optimized the geometry of DTLM molecule. Merz et al. evaluated the performance of an assortment of density functionals on the estimation of bond distances, bond angles and hydrogen bond energies, among other properties [21]. As a result, B1B95 was placed in the first position of the ranking (or second depending on the basis set) and B3LYP was classified as the sixth one (or the fifth). M06-2X density functional has been included to deal with intermolecular interactions with a high contribution of dispersion forces [22]. In addition, for comparative purposes, *ab initio* Moller-Plesset second order perturbation theory (MP2) [23] has been used for DTLM geometry optimization as well. All these methods implemented the Pople's 6-31 G (d) and 6-31+G(d) basis sets [24, 25] and Dunning's cc-pVDZ [26] and aug-cc-pVTZ [27] ones. The cc-pVTZ-DK was used to optimize the geometry of the metal complexes [28]. GDIIS algorithm [29, 30] was used for optimization along with tight convergence criteria. As for the self consistent field (SCF), tight convergence criterion was also imposed. The integral grid for two electron integrals and their derivatives was pruned to 99 radial shells and 590 angular points (integral=ultrafine). No symmetry restrictions were considered (Nosym).

As regards the title cluster, B1B95 functional was used to optimize its geometry involving Pople's 6-31 G(d) and 6-31+G (d) basis sets. GDIIS algorithm was considered in this process. Tight convergence criteria were imposed to the local minimum geometry and for SCF. The integral grid was pruned to ultrafine. The symmetry of the system, C_i, was taken into account.

The local minima for both free DTLM molecule and title cluster were assessed for all real vibrational frequencies but for DTLM molecule with MP2/aug-cc-pVTZ where computational difficulties prevent us to get the vibrational spectrum.

The evaluation of the energy for the intermolecular constants was accomplished within the approaches B1B95/ 6-31 G(d) and B1B95/6-31+G(d)//B1B95/6-31 G(d). Boyd and Bernardi's counterpoise correction was considered to account for the basis set superposition error (BSSE) [31, 32]. Difficulties arising from SCF impelled us to use the

Table 3 RMSs for the geometrical parameters of the cluster

RMS[a]	B1B95/6-31 G*	B1B95/6-31+G*	M06-2X/6-31 G(d)
X-Y, Å (D)	0.007	0.008	0.010
X-Y, Å (D+F)	0.009	0.009	0.011
X-Y, Å (D+F+I)	0.024	0.030	0.018
X-Y-Z, ° (D)	0.7	0.7	0.7
X-Y-Z, ° (D+F)	0.7	0.7	0.7

[a] X, Y, Z: Non-hydrogen atoms. (D): DTLM intramolecular parameters. (D+F): DTLM and DMF intramolecular parameters. (D+F+I): DTLM and DMF intramolecular distances and N···N and O···N intermolecular distances.

Table 4 Intermolecular geometrical data for the cluster

Parameters	Experimental	B1B95/6-31 G*	B1B95/6-31+G*	M06-2X/6-31 G(d)
N1···N8, Å	2.916(3)	3.008	3.033	2.962
N3···O1S, Å	2.835(3)	2.824	2.843	2.791
H1···N8, Å	2.05(1)	1.98	2.01	1.93
H3···O1S, Å	1.95(1)	1.79	1.81	1.75
N1-H1···N8, °	179(5)	177	177	178
N3-H3···O1S, °	177(3)	180	179	179

standard convergence criteria in energy calculations when BSSE correction was involved. Likewise, the integral grid was set to the standard values for Gaussian03 (integral=finegrid). It was checked that those modifications did not affect seriously the value of the energy respect to the calculations considering tight conditions.

The topology of the target systems has been analyzed within Bader's Atoms in molecule theory [33] by using AIM2000 code [34]. Popelier's eight criteria to characterize hydrogen bonding were used [35]: topology, ρ: electronic density [0.002-0.04]au, ε: ellipticity, $\nabla^2\rho$: Laplacian [0.02,0.15]au, Δr_T: mutual interpenetration, Δq: decrease of electronic charge in hydrogen atom, ΔH: destabilization of the hydrogen atom, $\Delta\mu$: decrease of hydrogen atom dipole moment, ΔV: decrease of hydrogen atom volume.

Discussion

Crystallographic part

The structure and numbering scheme of the H-bonded dimeric unit is shown in Fig. 1. Geometrical data are given in Table 1.

The C-S bond lengths are short if compared with typical thioureas (1.68-1.70 Å) [36] and clearly indicate the higher stability of the dithione form over the thiol ones. The C4-S4 bond lengths are similar to those observed in some pyrimidinethiones [37–39]; however, C2-S2 bond lengths are shorter than those found in the structures of 7-amino-quinazoline-2,4(1H,3H)-dithione [39] and 6-amino-2-thio-uracil [40]. The remaining bond distances are slightly shorter than those reported for some pteridinediones [41]. The maximum deviation from the best least-squares plane for pyridine ring is 0.031(2) Å for atom C(4) and for pyrazine, -0.025(2) Å for atom C(1A). The two rings of the pteridine system are nearly coplanar (acute dihedral angle 4.5°).

The crystal structure of the title compound is displayed in Fig. 2. The analysis of the hydrogen-bonds in the crystal structure indicates that neither an analogous dimerization pattern C=O···H-N, as found in related pteridinediones [9], nor the extensive N-H···S hydrogen bonds network found in the structure of 7-amino-quinazoline-2,4(1H,3H)-dithione [39] take place here, the preferential acceptor properties of the oxygen or nitrogen atoms over the sulfur ones being clearly shown. Thus, the pteridines are dimerized via N(1)-H···N8 (1-x, 1-y, 1-z) H-bonds, each one anchoring a DMF molecule by N3-H···O1S (x, 2-y, -½+z) hydrogen bonds

Table 5 Topology of the studied systems, M06-2X/6-31 G(d). A.u. for all parameters but for Δr_T (Å) and ΔH (kcal/mol)

Note: ′: Adjacent molecule in the same layer. ″: Adjacent molecule in the neighbor layer. ρ: electronic density [0.002-0.04]au, ε: ellipticity, $\nabla^2\rho$: Laplacian [0.02,0.15]au, Δr_T: mutual interpenetration, Δq: decrease of electronic charge in hydrogen atom, ΔH: destabilization of the hydrogen atom, $\Delta\mu$: decrease of hydrogen atom dipole moment, ΔV: decrease of hydrogen atom volume

Contact	$\rho(\varepsilon)$	$\nabla^2\rho$	Δr_T	Δq	ΔH	$\Delta\mu$	ΔV
Ligand intralayer							
N8′···H1-N1	0.034(0.037)	0.094	−0.953	0.034	0.011	−0.027	−9.563
O1S···H3-N3	0.044(0.022)	0.151	−0.998	0.071	0.029	−0.050	−12.025
S4···H1S-C1S	0.006(0.121)	0.019	−0.072	0.050	0.009	0.007	−1.670
S2···H7′-C7′	0.010(0.023)	0.030	−0.189	0.043	0.009	0.018	−3.806
Ligand interlayer							
S2···N1″	0.006(0.450)	0.018					
S2···N8″	0.009(0.349)	0.025					
O1S···N3″	0.010(1.299)	0.035					
Pd-complexes							
N5···Pd (A)	0.095(0.011)	0.502					
N1···Pd (B)	0.100(0.092)	0.533					
S4···Pd (A)	0.070(0.033)	0.199					
S2···Pd (B)	0.061(0.041)	0.165					
S2···H6′-C6′ (A)	0.011(0.264)	0.049	−0.736	0.021	0.009	−0.008	−3.983

(see Fig. 1); the dimers are roughly planar and are placed into two different set of planes -both containing the [−1,0,2] direction- mutually angled by 77.8°; the distance between two adjacent planes in each set is *ca.* 3.4 Å. Nevertheless, the analysis of short ring interactions shows too large slippage values (up to 3.8 Å) for aromatic rings from contiguous planes [42], the overlap between the π-electronic densities being consequently quite small.

A more detailed study suggests the existence of σ-π interactions (see Fig. 3) between the S2 and S4 sulfur atoms and pyrazine rings from adjacent molecules: C2-S2⋯Cg (pyrazine 1-x, 2-y, 2-z) 111.4(1)°, S2⋯Cg 3.675(2) Å; C4-S4⋯Cg(pyrazine x, 1+y, z) 81.0(1)°, S4⋯Cg 3.681(2) Å. Also, an additional interaction between the DMF carbonyl group and the pyrimidine ring is observed: C1S-O1S⋯Cg (pyrimidine x, 1-y, ½+z) 89.6(2)°, O1S⋯Cg 3.288(3) Å.

Structural modelization: free ligand

The analysis of both the molecular and supramolecular structure of the DTLM ligand is performed in three stages. Thus, the first stage concerns the study of its molecular geometry which allows doing an adequate benchmark of theoretical approaches in order to choose the most appropriate ones to simulate the supramolecular structure (Table 2).

The results were compared to the XRD experimental data for non-hydrogen atoms. The Table 2 supports the RMSs for each calculation method and kind of parameter, bond distances and angles. It is noteworthy that the molecular structure is predicted as planar for all methods.

As for the bond distances, it is observed that B1B95 performs better than B3LYP and MP2 irrespective of the basis set. The M06-2X/6-31 G(d) approximation yields intermediate values between the other two density functionals. The RMS values for B1B95 are within the typical experimental error if it is considered as 3σ. The values of the RMSs for bond distances are not significantly affected by the size and type of basis set. Only aug-cc-pVTZ exhibits slightly better values than the other basis sets for B3LYP and MP2, which are close to those ones yielded by B1B95.

The RMSs for the bond angles yield no significant differences for the density functionals irrespective of the basis set. As regards MP2 the results are slightly higher than for MP2/aug-cc-pVTZ. Anyway, the values are one order of magnitude higher than the experimental error.

Since the experimental crystal structure of the ligand involves layers of DTLM and formamide subunits, the second stage concerns the study of the intermolecular interactions stabilizing them inside the layers and the subsequent modifications in the molecular structures. Thus, a dimeric adduct is built up (Fig. 1) and its geometry
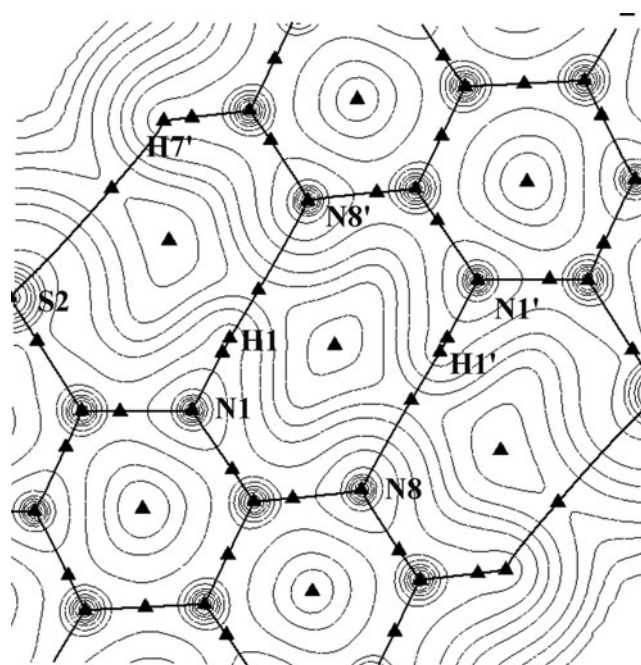

Fig. 4 Electronic density contours and topology for the intralayer N-H⋯N and C-H⋯S

simulated by using the B1B95/6-31 G(d), B1B95/6-31+G (d) and M06-2X/6-31 G(d) approaches provided they yield the best results for the DTLM monomer. The ratio results/ computational effort does not justify the use of Dunning basis sets. The selection of the 6-31+G(d) is interesting since the addition of a diffuse function can be *a priori* more appropriate to account for intermolecular contacts.

As concerns the RMSs for the [DTLM-DMF]₂ title cluster, it is observed in Table 3 that the bond distances for the DTLM molecular structure (X-Y(D)) have not been seriously affected for the cluster environment. Likewise, the
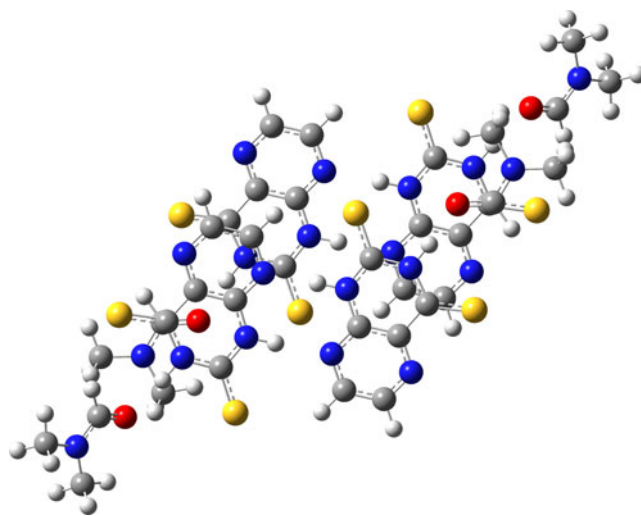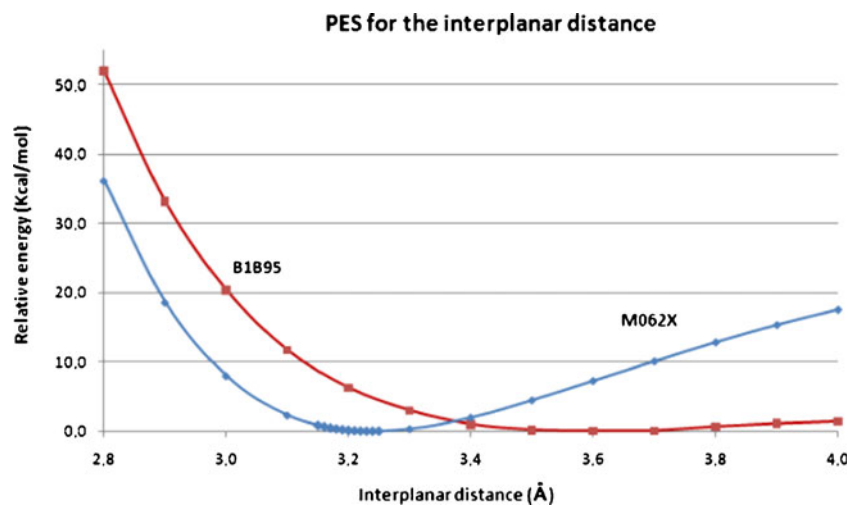

Fig. 5 Perpendicular view of the [DTLM-DMF]₄ cluster

Fig. 6 Test of the interplanar distance for [DTLM-DMF]$_4$. B1B95/6-31 G(d) and M06-2X/6-31 G(d) approaches

consideration of the dimethylformamide units does not disturb the agreement among calculated and experimental bond distances. On the other hand, the intramolecular bond angles for DTLM are significantly improved when the system is inserted in the cluster. In this case, the RMSs values are within or close to the typical experimental error, intended as 3σ. Focusing on concrete parameters some interesting results occur. Thus, the formation of the dimer DTLM-formamide involves an enlargement of both r(N3-H3) and r(C1S-O1S) bonds (0.024 Å and 0.012 Å respectively, M06-2X/6-31 G(d)) and a shortening of r(N3-C4) and r(N1S-C1S) (0.010 Å and 0.013 Å respectively, M06-2X/6-31 G(d)) which is coherent with a link by hydrogen bond. As for the DTLM$_2$ dimer, an increase of the bond r(N1-H1) as well as a shortening of the bond r(C2-N1) are observed (0.018 Å and 0.005 Å respectively, M06-2X/6-31 G(d)) which also points to hydrogen bonding. Likewise, the distance r(C2-S2) in DTLM$_2$ and r(C4-S4) in DTLM-formamide are increased suggesting some sort of intermolecular interaction through the S2 atom. This feature will be confirmed later on by the topological analysis.

The formation of the dimeric adduct [DTLM-DMF]$_2$ does not involve significant changes in the parameters commented previously but an enlargement for the intermolecular distances r(N···N) and r(N-H···N) (0.005 Å and 0.006 Å respectively is observed, M06-2X/6-31 G(d)) which suggests a weakening of the corresponding hydrogen bond. The consideration of the intermolecular distances in the estimation of the RMSs reveals that the best results are obtained for M06-2X/6-31 G(d) due to a better simulation of the r(N···N) distance (Table 4).

The intralayer intermolecular interaction stabilizing the dimeric adduct is characterized by the topology of the system within Bader's atoms in molecule theory (AIM) [33] and by using Popelier's eight criteria for hydrogen bonding [35]. Table 5 supports the relevant topological data. It can

be observed that both contacts C=O···H-N and N···H-N fulfill the eight criteria. Focusing on the most important ones, a critical point appears between the involved atoms (Fig. 4); the electronic density (ρ) is in the range [0.002, 0.04] a.u. and the positive value of the Laplacian $\nabla^2\rho$ characterize the critical points as minima, i.e., the electronic charge is deplete toward the atoms as expected in the so-called *closed shell interactions* [33]. Finally, both contacts reflect mutual interpenetration (Δr_T) which is sufficient criterion to address them as hydrogen bonds.

The S···H-C contacts fulfill all the criteria but the decrease of the dipolar moment ($\Delta\mu$). Also, S4 does not satisfy the mutual interpenetration criterion, the whole value is correct due to the H1S. Anyway, the low value of the electronic density characterizes these contacts as closer to van der Waals interactions than to hydrogen bonds [43].

In order to analyze the interlayer contacts the chemical model has been improved by superimposition of two dimeric adducts (Fig. 5), it will be addressed as [DTLM-DMF]$_4$. First, it has to be noted that the [DTLM-DMF]$_4$

Table 6 Interlayer geometrical data (Å)

Parameter	Distance to centroid (Å)	Proximal atoms	Distance (Å)
C1S=O···centroid	2.84 (3.29)	O1S…N3″	3.02 (3.55)
		O1S…C2″	3.10 (3.35)
		O1S…N1″	3.19 (3.30)
		O1S…C1A″	3.23 (3.56)
S2···centroid	4.14 (3.68)	S2…N8″	3.34 (3.46)
		S2…N1″	3.57 (3.85)
S4···centroid	3.87 (3.68)	S4…C6″	3.58 (3.73)
		S4…N5″	3.56 (3.84)

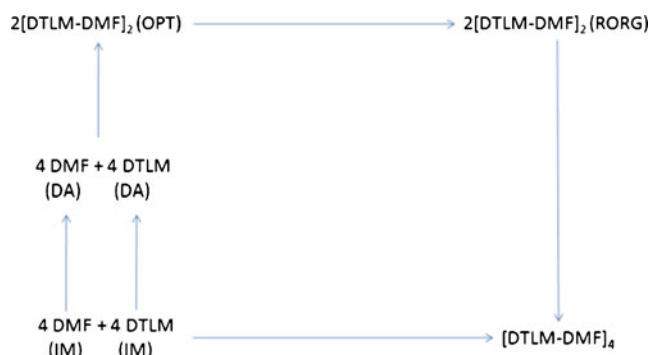Note: ″Atoms in the parallel layer. Experimental data in parentheses

Fig. 7 Scheme for the formation of the [DTLM-DMF]$_4$ cluster

cluster involves σ-π interlayer-intermolecular interactions with a high contribution of dispersion forces. Thus, the B1B95 density functional fails completely to produce a good model (Fig. 6). Only the M06-2X can be used provided it is optimized to account for dispersion.

As regards to the geometry of the [DTLM-DMF]$_4$ cluster, it is observed a shortening of the r(C=O···H-N) distance (0.061 Å) and an enlargement of the C=O one (0.008 Å), i.e., this hydrogen bond is strengthened. On the other side, the distances r(N-H···N) and r(N···N) are increased (0.044 Å and 0.034 Å, respectively).

As for the interlayer contacts some significant geometrical data are exhibited in Table 6. It is observed that the theoretical model underestimates the C=O···centroid distance and overestimates the S···centroid ones. These values along with the interatomic distances reveal some possibilities of σ-π

intermolecular contacts for O1S and S2 but as far as S4 is concerned the situation is doubtful. In order to be more conclusive an analysis of the topology is performed. Table 5 reveals significant critical points for the interlayer contacts O1S···N3″, S2···N1″ and S2···N8″. The values of ρ along with the positive value of the Laplacian indicate that these interactions are closed shell ones of the van der Waals type. It is significant the high value of the ellipticity which means the electronic charge is more extended than in the already commented intralayer interactions which makes sense if dispersion forces are involved.

Finally, a thermodynamical analysis is done for the formation of both the [DTLM-DMF]$_2$ and the [DTLM-DMF]$_4$ according to the scheme appearing in Fig. 7. The data provided in Table 7 shows that layer adduct [DTLM-DMF]$_2$ that accounts for the intralayer interactions is highly stable −21.8 kcal·mol^{-1} and −32.2 kcal·mol^{-1} for B1B95 and M06-2X density functionals implementing 6-31 G(d) basis set and BSSE. It is also observed the cooperative effect is small < 1 kcal·mol^{-1}. However, the consideration of the Gibbs' free energy shows the process for [DTLM-DMF]$_2$ is not spontaneous within the B1B95 approaches. The M06-2X/6-31 G(d) approximation produces a negative value of ΔG that turns to ~0 kcal·mol^{-1} if the BSSE is taken into account. The differences in the ΔG between the two functionals is the same as in the ΔH, ~10 kcal·mol^{-1}, then the better performance of M06-2X with respect to B1B95 is due to more adequate simulation of the binding energy which is related to the feature commented previously related to a best fit among the experimental and theoretical values of

Table 7 Binding enthalpy and Gibbs' free energy for [DTLM-DMF]$_2$ (kcal·mol^{-1})

ΔH (kcal·mol^{-1})	ΔH$_{reorg}$ (DTLM)	ΔH$_{reorg}$ (DMF)	ΔH$_{TC}$ ([DTLM-DMF])	ΔH$_{TC}$ ([DTLM]$_2$)	2ΔH$_{TC}$([DTLM-DMF])+ ΔH$_{TC}$([DTLM]$_2$)	ΔH$_{OC}$ ([DTLM-DMF]$_2$)	ΔH$_T$ ([DTLM-MF]$_2$)	ΔH$_{CE}$
B1B95/6-31 G*	0.62	0.34	−8.00	−11.95	−27.95	−26.46	−28.38	−0.43
B1B95/6-31 G*+B	0.62	0.34	−5.93	−9.52	−21.37	−19.87	−21.79	−0.41
B1B95/6-31+G*	0.58	0.20	−6.76	−9.58	−23.10	−22.39	−23.96	−0.86
B1B95/6-31+G*+B	0.58	0.20	−6.35	−8.39	−21.08	−20.01	−21.58	−0.50
M06-2x/6-31 G*	0.68	0.36	−10.78	−16.63	−38.19	−36.44	−38.52	−0.32
M06-2x/6-31 G*+B	0.68	0.36	−8.79	−14.24	−31.83	−30.07	−32.15	−0.32
ΔG (kcal·mol^{-1})	ΔG$_{reorg}$ (DTLM)	ΔG$_{reorg}$ (DMF)	ΔG$_{TC}$ ([DTLM-DMF])	ΔG$_{TC}$ ([DTLM]$_2$)	2ΔG$_{TC}$([DTLM-DMF])+ ΔG$_{TC}$([DTLM]$_2$)	ΔG$_{OC}$ ([DTLM-DMF]$_2$)	ΔG$_T$ ([DTLM-MF]$_2$)	ΔG$_{CE}$
B1B95/6-31 G*	0.80	0.23	1.19	−0.64	1.75	2.55	0.48	−1.26
B1B95/6-31 G*+B	0.80	0.23	3.26	1.80	8.32	9.15	7.08	−1.25
B1B95/6-31+G*	0.77	0.05	4.37	1.41	10.15	12.85	11.21	−1.06
B1B95/6-31+G*+B	0.77	0.05	4.79	2.60	12.17	15.22	13.58	−1.41
M06-2x/6-31 G*	0.91	0.26	−1.35	−5.05	−7.74	−6.31	−8.65	−0.91
M06-2x/6-31 G*+B	0.91	0.26	0.64	−2.66	−1.37	0.06	−2.29	−0.91

*B: BSSE: Basis set superposition error (minus sign means stabilization enthalpy); ΔH$_{reorg}$(DTLM): Reorganization enthalpy for DTLM; ΔH$_{reorg}$(DMF): Reorganization enthalpy for DMF; ΔH$_{TC}$([DTLM-DMF]): Stabilization enthalpy for the dimer DTLM-DMF with the structure in the title cluster (TC); ΔH$_{TC}$([DTLM]$_2$): Stabilization enthalpy for the dimer DTLM-DTLM with the structure in the cluster (TC); 2ΔH$_{TC}$([DTLM-DMF])+ΔH$_{TC}$([DTLM]$_2$): Sum of twice the stabilization enthalpy of the DTLM-DMF set plus the stabilization enthalpy for DTLM-DTLM; ΔH$_{OC}$([DTLM-DMF]$_2$): Stabilization enthalpy for the tetramer (only intermolecular contacts, OC); ΔH$_T$([DTLM-DMF]$_2$): Stabilization enthalpy for the tetramer plus reorganization energies; ΔH$_{CE}$: Cooperative effect

the r(N⋯N) and r(N-H⋯N) distances for the N-H⋯N hydrogen bond.

Anyway, the previous values of ΔG suggest it is necessary to improve the chemical model to obtain negative values of this parameter and to justify the formation of the DTLM-DMF crystal. The thermodynamical data for the cluster [DTLM-DMF]$_4$ is reported in Table 8. A negative value of ΔG is observed for the global process of formation of [DTLM-DMF]$_4$ the monomers (-4.2 kcal·mol^{-1}) including the BSSE. It is also noteworthy the high stabilization enthalpy for the interlayer intermolecular interactions, -31.8 kcal, revealing the σ-π interaction as a significant contribution to crystal total energy. The reorganization energy does not play an important role in the formation process.

Structural modelization: coordination modes

The next step of the present research consists of the study of the bonding capabilities of the DTLM ligand in order to form metal complexes. Thus, two options have been analyzed which involve a Pd atom and the N1-deprotonated ligand (Fig. 8). Both structures are optimized by using B1B95 density functional, because it is the best for bond distances, implementing Douglas and Kroll version of the cc-pVTZ-DK basis set [28] to account more properly for the electronic configuration of the Pd atom.

As observed in Fig. 8, the N⋯Pd is longer in the A complex (~0.03 Å) with respect to the B one and the opposite occurs for the S⋯Pd distance which is longer in B (~0.07 Å). The coordination sphere for complex B is more strained than for A. The N5-Pd-S4 angle amounts to 84.7º building a five-membered ring with the palladium atom in complex A, but for complex B, a four membered ring is obtained showing a N1-Pd-S2 angle of 69.3º. These features are supported by the topology of the provided systems; it is observed that the contacts N⋯Pd and S⋯Pd for B show greater ellipticity than for B (Table 5). It is noteworthy that the positive value of the Laplacian for the commented interactions points out they are of the closed-shell (ionic) type. Additionally, an intermolecular interaction S⋯H-C fulfilling Popelier's eight criteria appears for complex A.
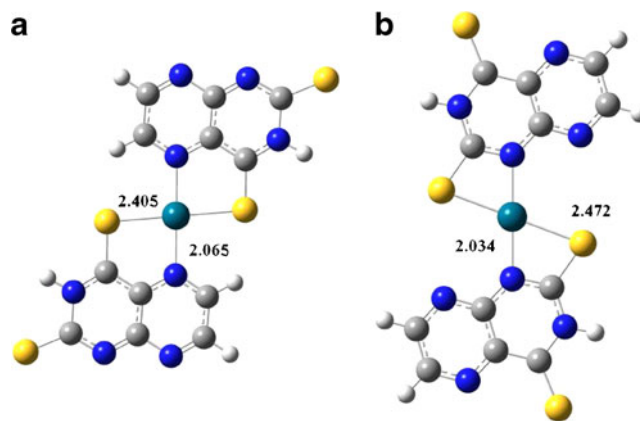

Fig. 8 Modelized N5-S4 (**a**) and N1-S2 (**b**) Pd complexes (M-L distances in Å)

In spite of these (*a priori*) arguments in favor of complex A, it turns out that complex B is 16.8 kcal·mol^{-1} more stable than A. The interaction enthalpy for complex A is -182.9 kcal·mol^{-1} and for B 199.8 kcal·mol^{-1}. ΔG exhibits negative for both complexes and the values are -159.2 kcal·mol^{-1} and 176.7 kcal·mol^{-1} respectively. Thus, the difference in the enthalpies is almost preserved in the free energy.

An explanation for this behavior can be found in the frontier molecular orbitals of the N1-deprotonated ligand. In general, as it is observed in Fig. 9 (A complex), some of the highest occupied molecular orbitals show a significant contribution of the S4 atom but a poor one from N5. On the other side, the contribution to such molecular orbitals of both S2 and N1 atoms (Fig. 9, B complex) is better balanced which can make complex B more favored energetically than A.
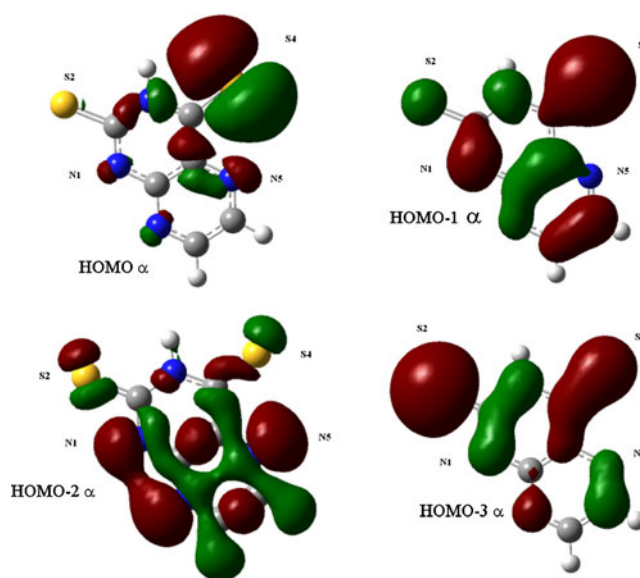
Table 8 Binding enthalpy and Gibbs' free energy for (kcal mol^{-1})

	[DTLM-DMF]$_2$ reorganization	[DTLM-DMF]$_4$ stabilization	Total from monomers
ΔH	1.59	−44.38	−114.07
ΔH+BSSE	1.83	−31.81	−88.29
ΔG	4.11	−25.58	−29.97
ΔG+BSSE	4.35	−13.00	−4.19

Fig. 9 Frontier molecular orbitals

References

1. Milstien S, Kapatos G, Levine RA, Shane B (2002) Chemistry and Biology of Pteridines and Folates, Proceedings of the 12th International Symposium on Pteridines and Folates. Kluwer, Boston, pp 60–139
2. Taylor EC (1993) Chemistry and Biology of Pteridines and Folates. In: Ayling JE, Nair MG, Baugh CM (eds) Advances in experimental medicine and biology, vol 338. Plenum, New York, pp 387–408
3. Prins LHA, Petzer JP, Malan SF (2009) Bioorg Med Chem 17:7523–7530
4. Ding Y, Girardet JL, Smith KL, Larson G, Prigaro B, Lai VCH, Zhong W, Wu JZ (2005) Bioorg Med Chem Lett 15:675–678
5. Gomtsyan A, Didomenico S, Chih-Hung L, Stewart AO, Bhagwat SS, Kowaluk A, Jarvis MF (2004) Bioorg Med Chem Lett 14:4165–4168
6. Schneider HJ, Pfleiderer W (1974) Chem Ber 107:3377–3394
7. Acuña-Cueva ER, Hueso-Ureña F, Jiménez-Pulido SB, Moreno-Carretero MN (2000) J Mol Model 6:433–437
8. Acuña-Cueva ER, Jiménez-Pulido SB, Moreno-Carretero MN (2002) J Mol Model 8:246–252
9. Acuña-Cueva ER, Faure R, Jiménez-Pulido SB, Moreno-Carretero MN, Peña-Ruiz T (2004) J Mol Struct 697:65–71
10. Pfleiderer W (1974) Comprehensive Heterocyclic Chemistry, vol. 3-2B. Pergamon, Oxford, p 309
11. Brown DJ (1988) In: Taylor EC (ed) Fused pyrimidines: pteridines, vol 24(3). Wiley, New York, pp 445–509
12. Sheldrick GM (2003) SADABS. University of Göttingen, Göttingen
13. Farrugia L (1999) J Appl Crystallogr 32:837–838
14. Sheldrick GM (1997) SHELXL97. University of Göttingen, Göttingen
15. Spek AL (2003) J Appl Crystallogr 36:7–13
16. Cambridge Crystallographic Data Centre MERCURY (2002). Cambridge, England
17. 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, Menucci G, 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 G, 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, Zakrzewsky 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, Ciolowski 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 E.01. Gaussian Inc, Wallingford, CT
18. Becke AD (1993) J Chem Phys 98:5648–5652
19. Becke AD (1996) J Chem Phys 104:1040–1046
20. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785–789
21. Riley KE, Op't Holt BT, Merz KM (2007) J Chem Theory Comput 3:407–433
22. Zhao Y, Truhlar DG (2008) Theor Chem Acc 120:215–241
23. Head-Gordon M, Pople JA, Frisch MJ (1988) Chem Phys Lett 153:503–506
24. Ditchfield R, Hehre WJ, Pople JA (1971) J Chem Phys 54:720–723
25. Clark T, Chandrasekhar J, Spitznagel GW, Schleyer PVR (1983) J Comput Chem 4:294–301
26. Woon DE, Dunning TH (1993) J Chem Phys 98:1358–1371
27. Kendall RA, Dunning TH, Harrison RJ (1992) J Chem Phys 96:6796–6806
28. Peterson KA, Figgen D, Dolg M, Stoll H (2007) J Chem Phys 126:124101–124112
29. Csaszar P, Pulay P (1984) J Mol Struct THEOCHEM 114:31–34
30. Farkas Ö, Schlegel HB (1999) J Chem Phys 111:10806–10814
31. Simon S, Duran M, Dannenberg JJ (1996) J Chem Phys 105:11024–11031
32. Boys SF, Bernardi F (1970) Mol Phys 19:553–566
33. Bader RF (1990) Atoms in molecules: a quantum theory. Oxford University Press, Oxford
34. Biegler-König F, Schönbohm J, Bayles D (2001) J Comput Chem 22:545–559
35. Koch U, Popelier PLA (1995) J Phys Chem 99:9747–9754
36. Allen FH, Watson DG, Brammer L, Orpen AG, Taylor R (2006) In: Prince E (ed) International Tables for Crystallography, vol C, ch 9.5. Springer, Heidelberg, pp 790–811
37. Takechi H, Kubo K, Takahashi H, Matsumoto T (2007) Acta Crystallogr E 63:o701–o703
38. Janik A, Olech A, Stasiewicz-Urban A, Stadnicka K (2009) Acta Crystallogr C 65:o70–o75
39. Yang KB, Lin LR, Huang RB, Zheng LS (2006) Acta Crystallogr E 62:o1938–o1940
40. Swaminathan S, Chacko KK (1978) Acta Crystallogr B 34:3108–3110
41. Hueso-Ureña F, Jiménez-Pulido SB, Moreno-Carretero MN, Quirós-Olozábal M, Salas-Peregrín JM (1998) Polyhedron 18:85–91
42. Janiak C (2000) J Chem Soc Dalton Trans 3885–3896
43. Varadwaj PR, Marques HM (2010) Theor Chem Acc 127:711–725

ORIGINAL PAPER

Four-component pharmacophore model for endomorphins toward μ opioid receptor subtypes

Yng-Ching Wu · Tim Jaglinski · Jin-Yuan Hsieh ·
Jia-Jyun Chiu · Tzen-Yuh Chiang · Chi-Chuan Hwang

Abstract In the present work, a series of simulation tools were used to determine structure-activity relationships for the endomorphins (EMs) and derive μ-pharmacophore models for these peptides. Potential lowest energy conformations were determined in vacuo by systematically varying the torsional angles of the Tyr^1-Pro^2 (ω_1) and Pro^2-Trp^3/Phe^3 (ω_2) as tuning parameters in AM1 calculations. These initial models were then exposed to aqueous conditions via molecular dynamics simulations. In aqueous solution, the simulations suggest that endomorphin conformers strongly favor the *trans/trans* pair of the ω_1/ω_2 amide bonds. From two-dimensional probability distributions of the ring-to-ring distances with respect to the pharmacophoric angles for EMs, a selectivity range of μ_1 is ca. 8.3~10.5 Å for endomorphin-2 and selectivity range of μ_2 is ca. 10.5~13.0 Å for endomorphin-1 were determined. Four-component μ-pharmacophore models are proposed for EMs and are compared to the previously published δ- and κ-pharmacophore models.

Introduction

It is well established that the δ, μ, and κ receptors are the primary locations where opiate drugs exert their influence [1–3]. It has been shown that the μ-selective agonists display the most potent antinociceptive activity, and further, of the known mammalian opioids, the endogenous peptides endomorphin-1 (EM1; Tyr-Pro-Trp-Phe-NH2) and endomorphin-2 (EM2; Tyr-Pro-Phe-Phe-NH2) exhibit the highest selectivity and binding affinity toward the μ-opioid receptor (MOR) [4, 5]. Hence, EM1 and EM2, as well as their derivatives, have been studied extensively to determine their structure-activity relationships (SAR) with the intent to design efficient analgesics with minimal side effects [6, 7].

Even though EM1 and EM2 have been targeted as model peptides, details of their three-dimensional structures and binding sites are still unclear. This is attributed to the fact that baseline solution and membrane-bound structural conformations of EM1 and EM2 were determined through various theoretical and experimental methods in different environments [8–13]. Not only do discrepancies exist among the currently accepted bioactive endomorphin conformation models, these models may not reflect the true in vivo conformation.

In opioid peptides, it is generally accepted that the most important pharmacophoric parameters are the distances from the protonated amine to the Tyr^1 aromatic ring, the distance from the protonated amine to the hydrophobic center and the distance from the Tyr^1 ring to the

Y.-C. Wu · T. Jaglinski · C.-C. Hwang (✉)
Biomedical Engineering Program, Department of Engineering Science, National Cheng Kung University,
Tainan 70101, Taiwan
e-mail: chchwang@mail.ncku.edu.tw

J.-Y. Hsieh
Computer Science Program, Department of Mechanical Engineering, Ming Hsin University of Science and Technology,
Hsinchu 30401, Taiwan

J.-J. Chiu
Biomolecular Network Program, Department of Electrical Engineering, National Cheng Kung University,
Tainan 70101, Taiwan

T.-Y. Chiang
Neuroscience Program, Department of Life Science,
National Cheng Kung University,
Tainan 70101, Taiwan

hydrophobic center [14]. However, EM1 and EM2 possess two hydrophobic centers, the Trp[3]/Phe[3] and Phe[4] aromatic rings situated at residue positions 3 and 4, and it has been pointed out that the N-terminal protonated amine, the Tyr residue in position 1 and both of the aromatic residues (Trp or Phe) in positions 3 and 4 can be considered pharmacophoric elements [15]. Hence, the usual understanding of the correlation of three-component 3D pharmacophore with bioactivity may not adequately predict the SAR of endomorphins (EMs).

Therefore, the present study attempts to investigate the SAR of EMs in vacuo and in aqueous solution using AM1 calculations and molecular dynamics (MD) simulations, respectively. The structural profiles of EM1 and EM2 in vacuo were examined by conducting a systematic conformer analysis using Spartan program. As a result, we obtained the $13 \times 13 = 169$ conformers of minimized energy for EM1 and EM2, respectively. Furthermore, in aqueous solution, the lowest energy conformers of EMs obtained from AM1 calculations were further analyzed using MD simulations. The results of these MD simulations suggest a MOR pharmacophore model which uses four-component (**N**, **A**, **B**, **C**) pharmacophoric elements rather than the usual three component system.

Methods

Method overview

Figure 1a and b illustrates the geometric structure, including the rotamers of the backbone and sidechains, for EM1 and EM2, respectively. The computational method used in the present work can be generally outlined as follows. Initial geometric endomorphin models were built using the commercial Spartan analysis package. Molecular mechanics (MM) calculations were used to obtain 169 lowest energy conformations in vacuo by varying the torsional angles of the Tyr[1]-Pro[2] (ω_1) and Pro[2]-Trp[3]/Phe[3] (ω_2) as tuning parameters. Using semi-empirical AM1 calculations, the 169 obtained conformations were further refined. Finally, MD simulations of the refined conformations in aqueous solution were conducted. From the aqueous conformations, the pharmacophore models were derived. For the present calculations, the ω_1 torsional angle was selected as the primary criterion to calculate the conformer profiles since the cis/trans isomerization occurs at the Tyr[1]-Pro[2] amide bonds. The ω_2 torsional angle was selected as reference gauge for calculating two-dimensional energy map.

In vacuo calculations

The initial endomorphin structures were constructed using the Spartan package from Wavefunction. The structural profiles

of EM1 and EM2 in vacuo ($\varepsilon = 1$) were then examined by conducting a systematic conformer analysis using Spartan program. At the first step, the torsional angles of ω_1 and ω_2 were both specified as 0°, and other rotamers were not restricted during the calculation process. The results of equilibrium conformer show that 1600 conformers for EM1 and 1444 conformers for EM2 were produced through MM calculations using the Merck molecular force field (MMFF), respectively [16, 17]. Only one of these conformers for EM1 and EM2, respectively, having the lowest energy was selected and then the equilibrium geometry by semi-empirical AM1 calculations was further calculated [18]. Subsequently, the torsional angles of ω_1 and ω_2 were mutually rotated by increments of 30°, and repeat the first step. Finally, we can obtain the $13 \times 13 = 169$ conformers of minimized energy for EM1 and EM2, respectively.

In aqueous solution calculations

MD simulations were performed using the GROMACS simulation program using the GROMOS96 (ffG43a1) force field [19, 20]. These simulations assume random atomic velocities consistent with a Maxwellian distribution. The atomic positions and velocities were integrated using the standard Verlet algorithm with a time step of 2 fs and SHAKE [21] to constrain all bond lengths. Simulations were performed at a constant temperature of T=300 K and a constant pressure of P=1 atm (the *NPT* ensemble) using the Berendsen coupling scheme [22] with a dimensionless time constant of 0.2. The particle mesh Ewald method (PME) [23, 24] was adopted to calculate the electrostatic forces, and the non-bonded potentials were truncated using a cutoff radius of 14 Å. The formation of a hydrogen bond between an atomic pair was set to occur if both the bond length (r_{HB}) ≤ 3.5 Å and the bond angle (θ_{HB}) ≤ 30° were satisfied. Coordinates and topologies from the lowest energy conformers of EMs obtained from AM1 calculations were converted using the PRODRG small-molecule topology generator program [25], and further analyzed in the GROMACS package. The EMs were solvated in boxes of SPC water [26] with a minimum distance of 10 Å from any peptide atom to the edge of the box. This resulted in approximately 1420 water molecules in a box of length 35.7 Å for EM1 and 1405 water molecules in a box of length 35.3 Å for EM2. Periodic boundary conditions were implemented.

Results

EM conformations in vacuo

Table 1 summarizes the obtained structural features for the EM1 and EM2 lowest energy conformations determined

Fig. 1 Schematic of the structural models of EMs. (**a**) EM1. (**b**) EM2. The pharmacophoric elements (red) include protonated nitrogen (**N**), centroid of phenolic group (**A**), and two centroid of hydrophobic groups (**B** and **C**). The key amide bonds are ω_1 and ω_2. The rotamers of the structural backbone and the sidechain are marked
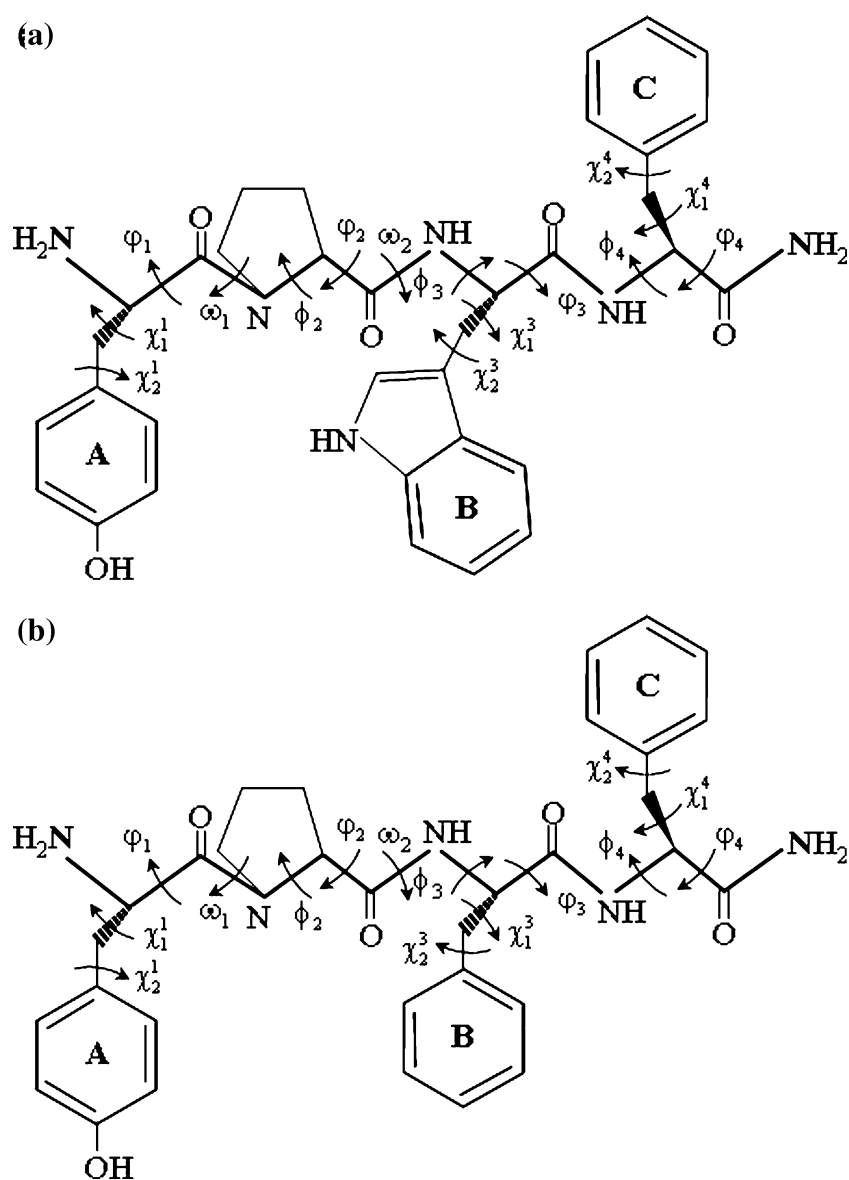
from the present quantum chemistry calculations in vacuo. Included are the amide bond rotamers (ω_1/ω_2), the end-to-end distance (r_{ee}), the probable pharmacophoric distances (**NA**, **NB**, **NC**, **AB**, **AC**, **BC**) and the side chain rotamer angles ($\chi_1^1, \chi_1^3, \chi_1^4$). Our results show that the lower minimized energy conformers of EMs are almost restricted to the conformers of *cis* (~ ±0°) and

trans (~ ±180°) isomerization regarding the ω_1 and ω_2 rotamers. Nevertheless, EMs bear the lowest energy conformers while both ω_1 and ω_2 rotamers lied on *trans* form. This can be seen from the smoothed, two-dimensional energy maps for EM1 and EM2 (shown as Fig. 2a, b, respectively) constructed from the 169 lowest energy conformations. The lowest energy conformers for

Table 1 Comparison of the lowest-energy conformers of EM1 and EM2 in vacuo using AM1 calculations

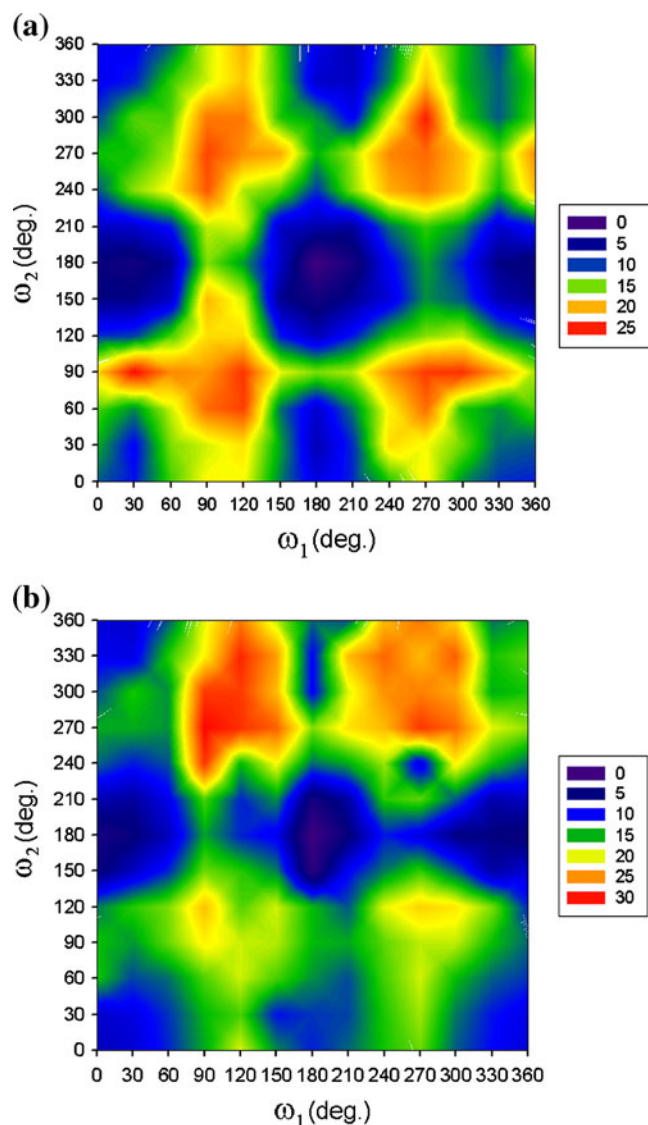| ligand	Amide bond rotamer (°)		End-to-end distance (Å)	N-to-ring distance (Å)			Ring-to-ring distance (Å)			Sidechain rotamer (°)		
	ω_1/ω_2		r_{ee}	NA	NB	NC	AB	AC	BC	χ_1^1	χ_1^3	χ_1^4
EM1	trans/trans	180/180	9.85	3.99	8.54	5.96	11.32	6.76	9.58	-56.52	-58.12	-56.77
EM2	trans/trans	180/180	7.76	5.10	10.40	11.63	7.34	8.90	5.44	-172.88	-157.88	-65.35

Fig. 2 Conformer energy maps of EMs in terms of the ω_1 and ω_2 torsional angles in vacuo from AM1 calculations. Torsional angles of ω_1 and ω_2 correspond to Tyr[1]-Pro[2] and Pro[2]-Trp[3]/Phe[3] amide bonds, respectively. (**a**) EM1. (**b**) EM2

Fig. 3 The lowest energy conformers of EMs in vacuo. (**a**) EM1. (**b**) EM2. Both are at the lowest energy state with three intramolecular HBs shown in yellow dot line

shown in Table 3, it is can be seen that the backbone dihedral angles of EMs reveal the conformers characterized by a 3→1 inverse γ-turn and a 4→2 regular γ-turn with respect to AM1 calculations.

EM conformations in aqueous solution

In regards to structural stability, time histories of the root-mean-square-deviation (RMSD) value of the EMs backbone atoms (Fig. 4a) and the end-to-end distances (r_{ee}) (Fig. 4b) show that stable structures are obtained for both EMs in 2 ns or less of simulation time. After the full

EM1 and EM2 in vacuo are depicted in Fig. 3. Three intramolecular hydrogen bonds (HBs) are predicted (dashed lines in Fig. 3) and summarized in Table 2. Both EMs possess Trp[3]:NH to Tyr[1]:CO (3→1) and Phe[4]:NH to Pro[2]:CO (4→2) intramolecular HBs. EM1 contains a third intramolecular hydrogen bond from the C-terminal NH2 group to the CO group of Trp[3]. EM2 also contains a third hydrogen bond from the OH group of the phenolic ring to the backbone CO group of the Phe[3] residue. It was calculated that the $\chi_1^1, \chi_1^3, \chi_1^4$ sidechain angles favor the *gauche* (−) (~ − 60°) and *trans* rotamers in the lowest energy conformers. For EM1, the ($\chi_1^1, \chi_1^3, \chi_1^4$) rotamers are (*gauche* (−), *gauche* (−), *gauche* (−)), and for EM2 the rotamers are (*trans*, *trans*, *gauche* (−)). Furthermore, as
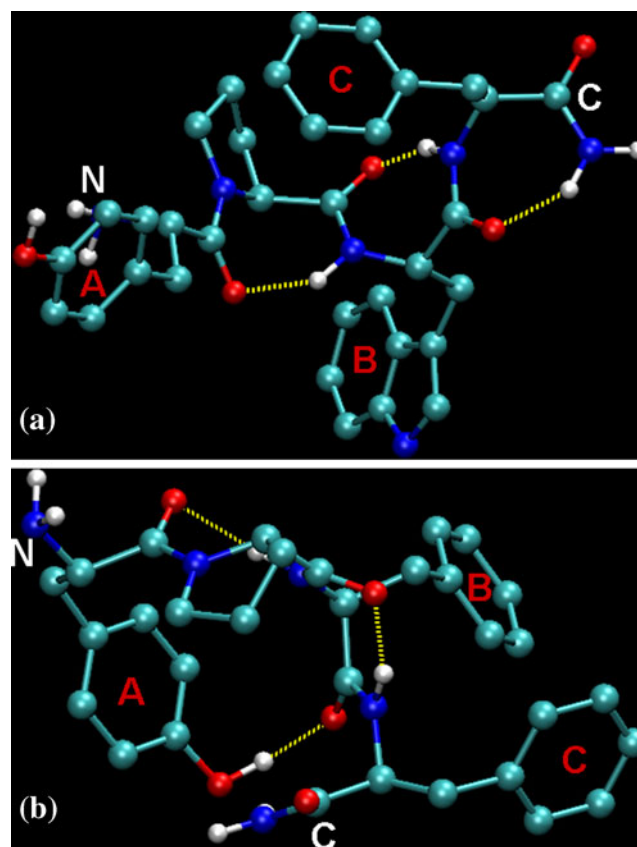
Table 2 Existence of hydrogen bonds in vacuo

Ligand	Donor residue → acceptor residue
EM1	3 → 1
	4 → 2
	C-terminal NH$_2$ group → CO group of Trp^3 residue
EM2	3 → 1
	4 → 2
	OH group of phenolic ring → CO group of Phe^3 residue

Table 3 Comparison of all the dihedral angles (°) of EMs in vacuo and in aqueous solution using AM1 calculations and MD simulations, respectively

		Backbone chain				Sidechain			
		φ		ψ		χ^1		χ^2	
		AM1	MD	AM1	MD	AM1	MD	AM1	MD
Tyr^1	EM1	–	–	72.89	-17.21	-56.52	-66.70	119.91	97.23
	EM2	–	–	93.37	103.31	-172.88	-174.43	74.23	76.27
Pro^2	EM1	-76.83	-45.43	56.89	113.50	–	–	–	–
	EM2	-79.83	-45.61	51.30	107.61	–	–	–	–
Trp^3	EM1	64.49	-84.90	-47.46	123.53	-58.12	-65.17	-76.78	-90.46
Phe^3	EM2	69.81	58.87	-57.99	-87.83	-157.88	-65.05	87.72	99.16
Phe^4	EM1	75.73	-136.22	-47.81	-44.18	-56.77	-66.27	110.41	106.08
	EM2	-92.71	-62.87	85.62	139.42	-65.35	-72.94	100.20	88.05

simulation time of 10 ns, only the ($\sim \pm 180°$) *trans* ω_1 conformation is obtained (Fig. 5c) and r_{ee} converges to average values of ca. 11.0 and 7.4 Å for EM1 and EM2, respectively.

Based upon the geometric criteria for the establishment of a hydrogen bond, EM1 has a $3 \rightarrow 1$ intramolecular hydrogen bond in aqueous solution (Fig. 5a). The average value of 3.5 Å between the Phe^4:NH and Pro^2:CO functional groups, i.e., $4 \rightarrow 2$ intramolecular hydrogen bond, suggests that a second hydrogen bond forms here as well, however, some uncertainty exists in this conclusion
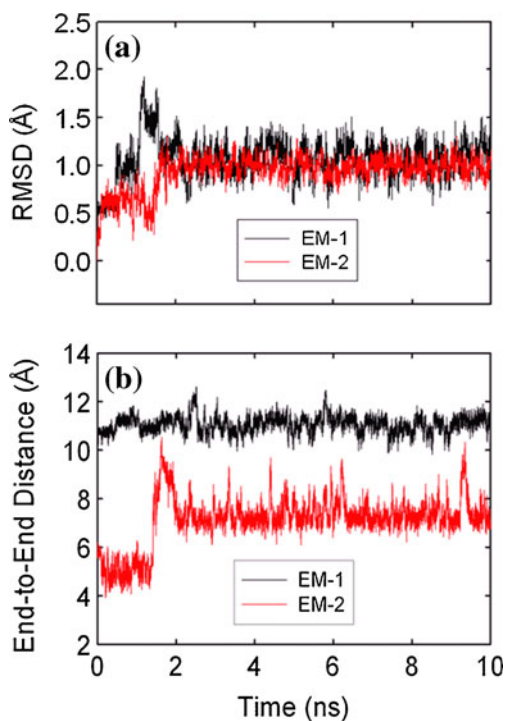

Fig. 4 The structural stability of EMs in aqueous solution. (**a**) The RMSD value of the backbone atoms corresponding to the initial structure obtained from AM1 calculations. (**b**) The variations of end-to-end distances. The distance is defined from the N-terminal nitrogen atom to C-terminal nitrogen atom
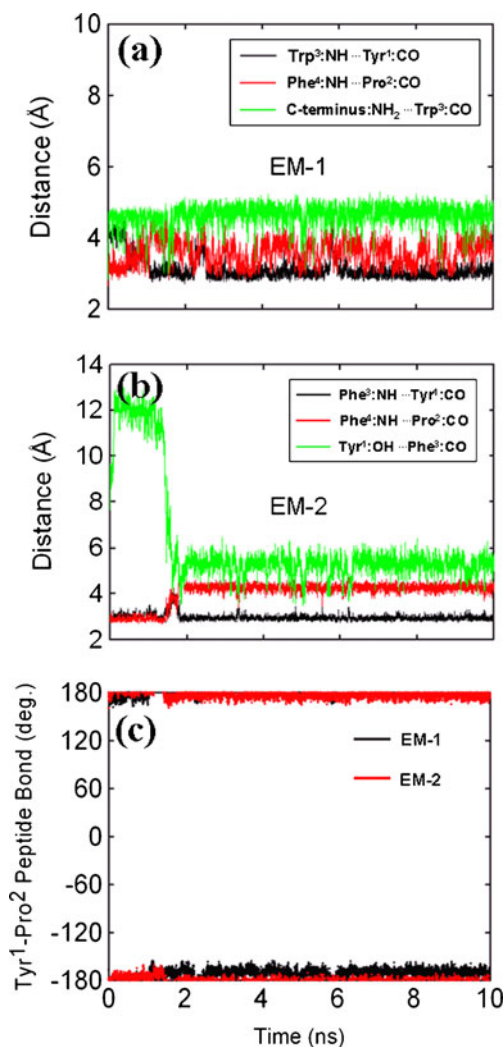

Fig. 5 MD simulations results for dynamic behavior of EMs in aqueous solution. (**a**) The variation of intramolecular HBs over time for EM1. (**b**) The variation of intramolecular HBs over time for EM2. (**c**) The time evolution of torsional angles of Tyr^1-Pro^2 peptide bond (ω_1) for Ems

Fig. 6 Time evolutions of side-chain orientations expressed in terms of χ_1 and χ_2 angles for selected residues for EM1. Horizontal lines indicate corresponding values of the lowest energy conformer for EM1 as computed by AM1 calculations. It can be seen that the χ_1^1, χ_1^3 and χ_1^4 angles all favor the *gauche* (−) rotamers in the stable structure
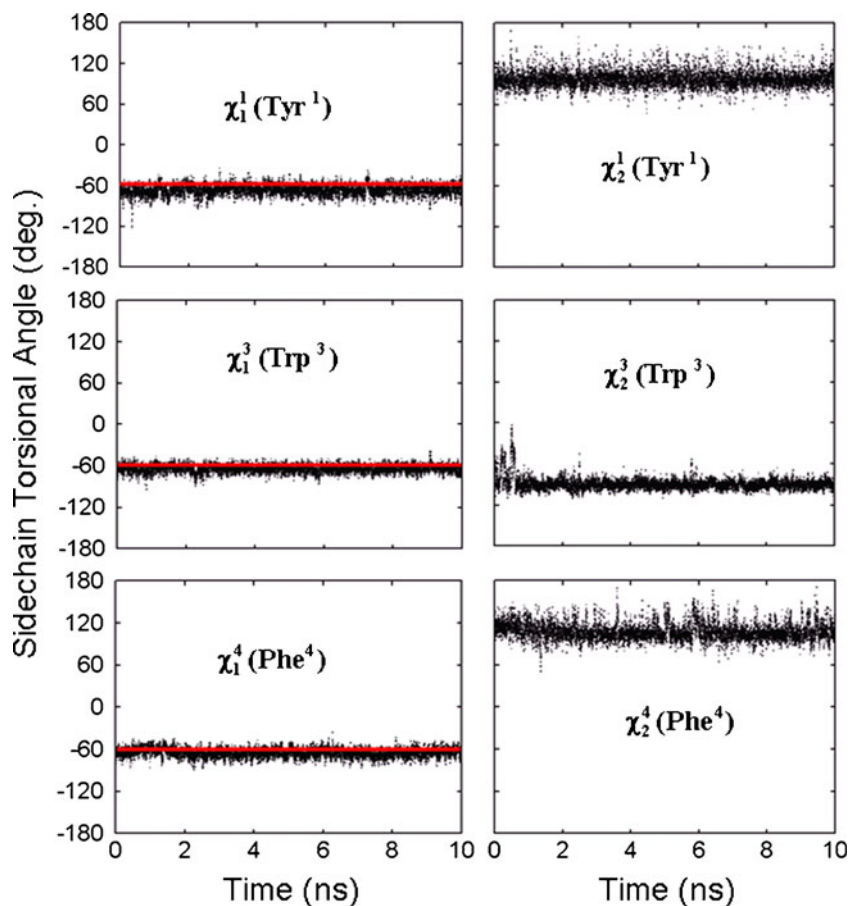
due to the calculation noise. In contrast to EM1, EM2 possesses only the 3 → 1 intramolecular hydrogen bond. The third intramolecular hydrogen bond predicted by AM1 calculations for EMs is absent in aqueous solution (Fig. 5b). Furthermore, no new HBs are formed during the MD simulations.

Table 3 also summarizes the obtained structural features for the EM1 and EM2 stable conformations determined from the present MD simulations in aqueous solution including the backbone and sidechain rotamers. For backbone dihedral angles, it can be seen that EM1 exhibits a 3 → 1 inverse γ-turn, as well as a 4 → 2 inverse γ-turn against regular γ-turn from AM1 calculations. In contrast, only a 3 → 1 inverse γ-turn is present in EM2. On the other hand, it is well-known that the χ_1 rotamers of aromatic rings (e.g., Tyr and Phe) play an essential role in determining the pharmacological activity. Consequently, the sidechain (χ_1^1, χ_1^3, χ_1^4) distributions of EM1 and EM2 in aqueous solution are analyzed as a function of time, as shown in Figs. 6 and 7, respectively. In Figs. 6 and 7, the horizontal lines indicate corresponding values of the lowest energy conformers for EMs as computed by AM1 calculations. In regards to structural stability, the results show that the (χ_1^1, χ_1^3, χ_1^4) rotamers prefer (*gauche* (−), *gauche* (−), *gauche* (−)) for EM1 in aqueous solution that are

similar to those observed in vacuo, whereas these rotamers prefer (*trans*, *gauche* (−), *gauche* (−)) for EM2 in aqueous solution that only the χ_1^3 rotamer dramatically varies against AM1 calculations.

Four-component μ-pharmacophore model

Since EMs bear three significant pharmacophoric rings (Tyr[1], Trp[3]/Phe[3], Phe[4]), four-component (**N**, **A**, **B**, **C**) pharmacophore models, based on the stable conformations in aqueous solution, are proposed for EM1 and EM2 (Fig. 8). Many of the possible distances between the endomorphin rings or functional groups fits within the framework of the two-ring models, suggesting that these EMs can bind to many receptors. For two-ring models of receptor selectivity, it has been suggested that aromatic ring separations of 10~13 Å are required for MOR activity [27, 28], 5.7~8.3 Å are required toward the δ opioid receptor (DOR) [29] and 5.0~5.4 Å are required for all three subtypes toward the κ opioid receptor (KOR) [30] immediately eliminating several distances. Further, since it is known from experimental evidence that EMs exhibit the highest selectivity and binding affinity toward the MOR, the calculations which predict that the average value of ring-to-ring distance **AC** (d(**AC**)) is ca. 12 Å for EM1 and
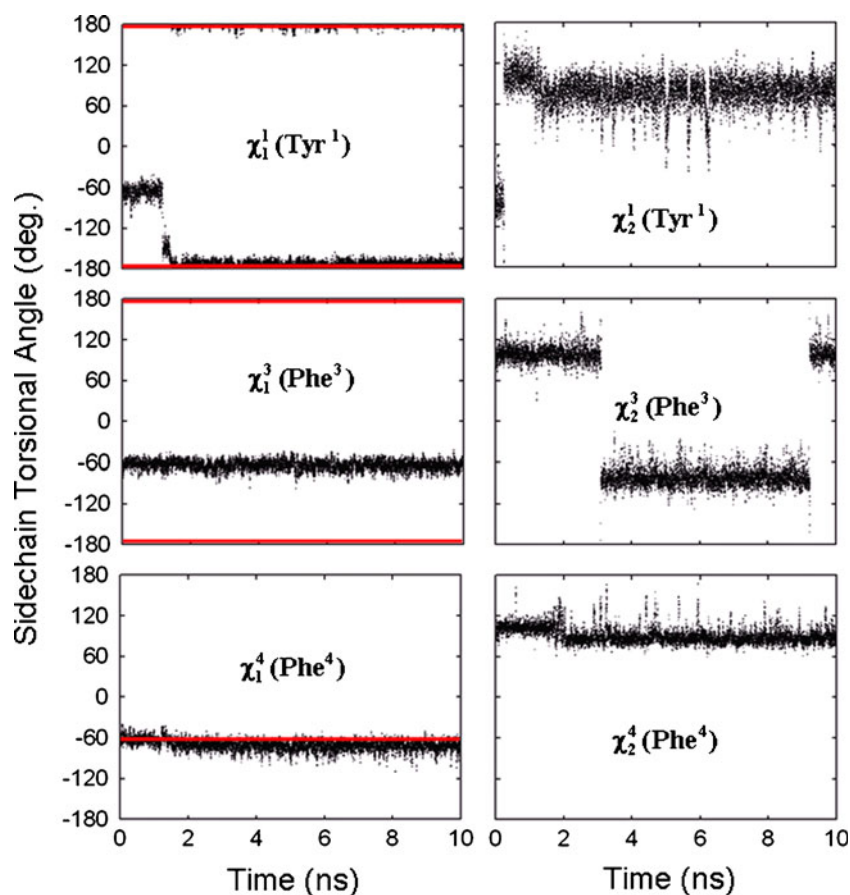
that ring-to-ring distance **AB** (d(**AB**)) is ca. 9.4 Å for EM2. However, to fully characterize the peptides, the four-component model proposed here is needed.

Discussion

As previously mentioned, EMs possess the highest selectivity and binding affinity toward the MOR, which has a selectivity range of 10–13 Å regarding two aromatic rings. However, there is biochemical and pharmacological evidence suggesting that MOR can be divided into the μ_1 and μ_2 subtypes [5, 31, 32]. In fact, experimental studies also further indicated that EM1 binds to μ_2 and that EM2 binds to μ_1 [5, 32]. Moreover, previous study has reported that overlap between DAMGO and EM1 conformers show the Tyr[1]-MePhe[4] of DAMGO and Tyr[1]-Phe[4] of EM1 can assume a similar orientation with a separation between the aromatic rings of ca. 12 Å [27], simultaneously, the results also suggested the Phe[4] residue of EM1 adopts a bioactive conformation at the receptor site and the Trp[3] residue appears as an additional site. On the other hand, in order to clarify the role of Phe[3] and Phe[4] of EM2 in receptor binding, prior research has designed and synthesized a series of analogs

in which these Phe residues were replaced by various amino acids [33]. The result indicated that Phe[3] residue has much stronger activity than Phe[4] residue. As mentioned above in connection with different binding characteristics toward the MOR subtypes for EMs, suggest that **N**, **A**, and **C** are the pertinent pharmacophoric elements of EM1, whereas **N**, **A**, and **B** are the pharmacophoric elements of EM2 (see the thick lines of Fig. 8).

The actual selectivity ranges between the aromatic rings for the MOR subtypes have not yet been determined, however, the size distributions of the d(**AC**) and d(**AB**) determined in our calculations could reveal possible binding selectivities for the different μ_1 and μ_2 subtypes. Lissajous construction of the 10 ns simulation time-histories results in the two-dimensional (2D) probability distributions of the distances **AB/C** (d(**AB**) for EM2 and d(**AC**) for EM1) against the endomorphin pharmacophoric angles shown in Fig. 9. In Fig. 9c, the designation for angle NB/CA stands for ∠NBA for EM2 and ∠NCA for EM1; similarly for the other angles. From the 2D distributions, the majority of the distribution range for d(**AC**) in EM1 is ca. 10.5~13.0 Å (blue), similarly, the range for d(**AB**) in EM2 is ca. 8.3~10.5 Å (red). Our finding suggests the selectivity ranges for μ_1 and μ_2 are ca. 8.3~10.5 Å and 10.5~13.0 Å, respectively.

(a)

d (**NA**)=4.03 ± 0.10   d (**AB**)=7.54 ± 0.74

d (**NB**)=4.14 ± 0.70   d (**AC**)=12.05 ± 0.73

d (**NC**)=9.72 ± 0.33   d (**BC**)=9.01 ± 0.29

(b)

d (**NA**)=5.17 ± 02   d (**AB**)=9.38 ± 0.34

d (**NB**)=9.75 ± 0.17   d (**AC**)=6.45 ± 0.48

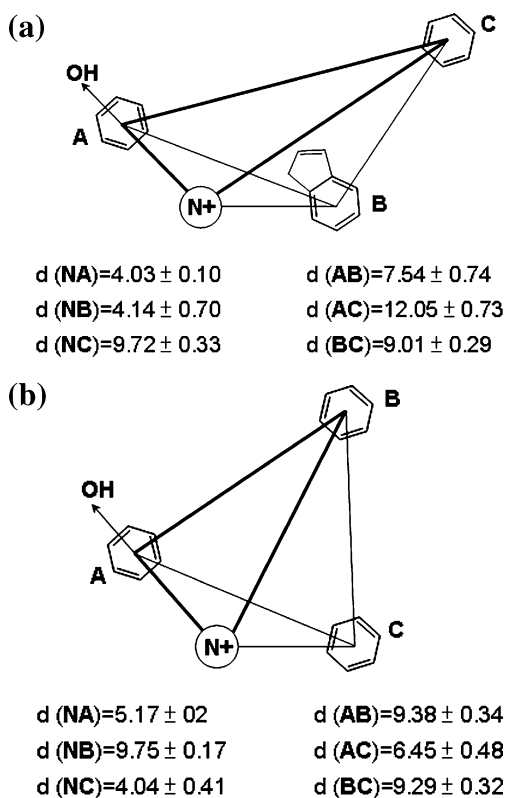d (**NC**)=4.04 ± 0.41   d (**BC**)=9.29 ± 0.32

Fig. 8 The four-component μ-pharmacophore models developed for (**a**) EM1. (**b**) EM2. The pairwise distances and the variance values between these four pharmacophore components are shown

In regards to the accuracy of the present simulations in aqueous solution, some connection with reality must be made to show that the bioactive aqueous solution conformations calculated in the present work are reasonable. First, the *trans* only ω_1 bioactive conformation predicted here is consistent with prior researches [10, 13, 34]. In addition, the intramolecular hydrogen bond binding and backbone structure behaviors predicted here are also in agreement with previous studies [10, 11, 34]. On the other hand, the results of sidechain orientations for structurally stable features of EMs indicate that the χ_1^1 rotamer of EM1 favors *gauche* (−), whereas EM2 favors *trans*. During the simulation time period of 10 ns, the initial χ_1^1 rotamer of EM2 adopts *gauche* (−) orientation, and after few ns (ca. 1.5 ns), it rapidly varies to *trans* orientation and remains the value to the end of simulation. This dynamic process also directly affects the interaction between the Tyr[1]:OH and Phe[3]:CO functional groups. As a result, this distance interaction rapidly drops to about 5 Å at around 1.5 ns (see Fig. 5b). However, these results are in agreement with the findings of previous studies which indicated the Tyr[1] prefers to accommodate both the *trans* and *gauche* (−) sidechain orientations [35]. Furthermore, in the present study, the *gauche* (−) orientations for both the χ_1^3 and χ_1^4 rotamers have also been suggested for EMs and their

analogues [36–38]. The main tri-rotamer-combinations (χ_1^1, χ_1^3, χ_1^4) of sidechain rotamers of the Tyr[1], Trp[3] and Phe[4] residues for EM1 and the Tyr[1], Phe[3] and Phe[4] residues for EM2 have been identified in χ_1 conformational space [36]. The considerable results indicated that four types of these combinations exist in the conformers of all types of EMs, i.e., (*gauche* (−), *gauche* (−), *gauche* (−)), (*trans*, *gauche* (−), *gauche* (−)), (*gauche* (−), *gauche* (−), *trans*), and (*trans*, *trans*, *gauche* (−)). Indeed, our results show that the tri-rotamer-combinations of EMs are in agreement with previous studies.

Despite these consistencies with previous work, some debate still exists about the detailed structure information and binding sites of these EMs. For example, peptide structures are determined experimentally using cryo-electron microscopy. This technique can only reveal conformation snapshots at temperatures much lower than in the human body and it is not clear how much the
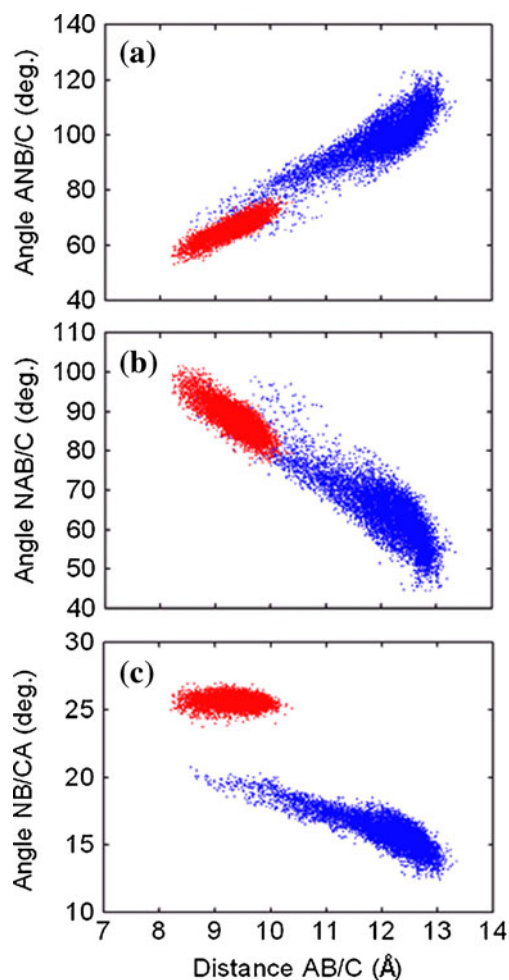

Fig. 9 2D probability distributions of the distances against the pharmacophoric angles for EMs. The distance **AC** for EM1 is shown in blue color, whereas distance **AB** for EM2 is shown in red color. The angle NB/CA stands for ∠NBA for EM2 and ∠NCA for EM1; similar for other angles

cryogenic conformations deviate from those in vivo. Our simulations can demonstrate this, though somewhat indirectly. Quantum chemistry calculations are conducted at 0 K in vacuo. In the present work, three hydrogen bonds, which effect ring distances, were predicted for both EM1 and EM2 in vacuo, however several of these bonds were not predicted in aqueous solution. Though not totally accurate, our quantum chemistry calculations and subsequent MD simulations do provide realistic estimates of in vivo conditions.

It is acknowledged that the models used herein do not completely capture the physical situation in vivo. For example, conformation changes incurred from interactions of the EMs in the locality of the receptors were not modeled. Inclusion of the receptors into the models, though warranted, was not sensible based on the computing power available. As for the flexibility of linear peptides, conformers other than the ones calculated here could exist in vivo. The lowest energy conformers determined from the quantum chemistry calculations only provide initial estimates of the peptide configurations. However, the MD simulations in aqueous do include thermal fluctuations, which are necessary to overcome energy barriers between possible lower energy states. Hence, the MD simulations explore a large energy space. From our analysis, gross changes to the overall structures do not occur through the course of the MD simulations suggesting that the structures determined from quantum chemistry calculations are close to their lowest energy configuration. However, for a sensible pharmacophore model, accurate determination of the distances between rings and other functional groups is required since the ranges for receptor selectivity are not widely separated. This is provided via MD simulations. The four-component model presented in the present work is intended to be a starting point as no model yet exists for these peptides. The distances presented here should be verified through future experimental research.

Conclusions

In this study, we have identified the SAR of EMs in vacuo and in aqueous solution using AM1 calculations and MD simulations, respectively. In vacuo, it was found that the most stable conformers for both the EM1 and EM2 endogenous opioid peptides contain three intramolecular hydrogen bonds. Furthermore, it is can be seen in Table 3, the backbone dihedral angles of EMs reveal the conformers characterized by a $3 \rightarrow 1$ inverse γ-turn and a $4 \rightarrow 2$ regular γ-turn, meanwhile, it has been shown that the $\chi_1^1, \chi_1^3, \chi_1^4$ sidechain angles favor the *gauche* (−) and *trans* rotamers in the lowest energy conformers. For EM1, the $(\chi_1^1, \chi_1^3, \chi_1^4)$ rotamers are (*gauche* (−), *gauche* (−), *gauche* (−)), and for

EM2 the rotamers are (*trans*, *trans*, *gauche* (−)). In aqueous solution, EM1 appears for the $3 \rightarrow 1$ and $4 \rightarrow 2$ intramolecular hydrogen bonds based on the conformational stability. By contrast, EM2 possesses only the $3 \rightarrow 1$ intramolecular hydrogen bond. For backbone conformational characteristics, EM1 exhibits a $3 \rightarrow 1$ inverse γ-turn, as well as a $4 \rightarrow 2$ inverse γ-turn against regular γ-turn from AM1 calculations, whereas EM2 only appears on a $3 \rightarrow 1$ inverse γ-turn. Furthermore, the results show that the $(\chi_1^1, \chi_1^3, \chi_1^4)$ rotamers prefer (*gauche* (−), *gauche* (−), *gauche* (−)) for EM1 that are similar to observed those in vacuo, whereas these rotamers prefer (*trans*, *gauche* (−), *gauche* (−)) for EM2 that only the χ_1^3 rotamer dramatically varies against AM1 calculations. Since the EMs contain three significant pharmacophoric rings (Tyr[1], Trp[3]/Phe[3], Phe[4]) in their sequences, the four-component (**N**, **A**, **B**, **C**) correlation is proposed. The results indicate that the **N**, **A**, and **C** are the main pharmacophoric elements for EM1, whereas **N**, **A**, and **B** are the pharmacophoric elements for EM2. Finally, from the 2D probability distributions of d(**AC**) for EM1 and d(**AB**) for EM2, our finding suggests the selectivity ranges for μ_1 and μ_2 subtypes are ca. 8.3~10.5 Å and 10.5 ~13.0 Å, respectively.

References

1. Clark JA, Liu L, Price M, Hersh B, Edelson M, Pasternak GW (1989) J Pharmacol Exp Ther 251:461–468
2. Harrison LM, Kastin AJ, Zadina JE (1998) Peptides 19:1603–1630
3. Kieffer BL (1999) Trends Pharmacol Sci 20:19–26
4. Zadina JE, Hackler L, Ge LJ, Kastin AJ (1997) Nature 386:499–502
5. Sakurada S, Zadina JE, Kastin AJ, Katsuyama S, Fujimura T, Murayama K, Yuki M, Ueda H, Sakurada T (1999) Eur J Pharmacol 372:25–30
6. Szatmári I, Biyashev D, Tömböly C, Tóth G, Mácsai M, Szabó G, Borsodi A, Lengyel I (2001) Biochem Biophys Res Commun 284:771–776
7. Okada Y, Fujita Y, Motoyama T, Tsuda Y, Yokoi T, Li T, Sasaki Y, Ambo A, Jinsmaa Y, Bryant SD, Lazarus LH (2003) Bioorg Med Chem 11:1983–1994
8. Fiori S, Renner C, Cramer J, Pegoraro S, Moroder L (1999) J Mol Biol 291:163–175
9. In Y, Minoura K, Ohishi H, Minakata H, Kamigauchi M, Sugiura M, Ishida T (2001) J Pept Res 58:399–412
10. In Y, Minoura K, Tomoo K, Sasaki Y, Lazarus LH, Okada Y, Ishida T (2005) FEBS J 272:5079–5097
11. Leitgeb B, Ötvös F, Tóth G (2003) Biopolymers 68:497–511
12. Podlogar BL, Paterlini MG, Ferguson DM, Leo GC, Demeter DA, Brown FK, Reitz AB (1998) FEBS Lett 439:13–20

13. Shao X, Gao Y, Zhu C, Liu X, Yao J, Cui Y, Wang R (2007) Bioorg Med Chem 15:3539–3547
14. Deschamps JR, Flippen-Anderson JL, George C (2002) Biopolymers 66:287–293
15. Gentilucci L, Tolomelli A (2004) Curr Top Med Chem 4:105–121
16. Halgren TA (1996) J Comput Chem 17:490–519
17. Halgren TA (1996) J Comput Chem 17:616–641
18. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902–3909
19. Lindahl E, Hess B, van der Spoel D (2001) J Mol Model 7:306–317
20. Scott WRP, Huenenberger P, Tironi I, Mark A, Billeter S, Fennen J, Torda A, Huber T, Krueger P, van Gunsteren W (1999) J Phys Chem A 103:3596–3607
21. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) J Comput Phys 23:327–341
22. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) J Chem Phys 81:3684–3690
23. Darden T, York D, Pedersen L (1993) J Chem Phys 98:10089–10092
24. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) J Chem Phys 103:8577–8593
25. Schuettelkopf AW, van Aalten DMF (2004) Acta Crystallogr D Biol Crystallogr 60:1355–1363
26. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) In: Pullman B (ed) Intermolecular forces. Reidel, Dordrecht, pp 331–342
27. Paterlini MG, Avitabile F, Ostrowski BG, Ferguson DM, Porthogese PS (2000) Biophys J 78:590–599
28. Yamazaki T, Pröbsti A, Schiller PW, Goodman M (1991) Int J Pept Protein Res 37:364–381
29. Shenderovich MD, Liao S, Qian X, Hruby VJ (2000) Biopolymers 53:565–580
30. Wu YC, Lin JS, Hwang CC (2007) J Phys Chem B 111:7377–7383
31. Pasternak GW, Wood PL (1986) Life Sci 38:1888–1898
32. Sakurada S, Hayashi T, Yuhki M, Fujimura T, Murayama K, Yonezawa A, Sakurada C, Takeshita M, Zadina JE, Kastin AJ, Sakurada T (2000) Brain Res 881:1–8
33. Honda T, Shirasu N, Isozaki K, Kawano M, Shigehiro D, Chuman Y, Fujita T, Nose T, Shimohigashi Y (2007) Bioorg Med Chem 15:3883–3888
34. Leitgeb B, Szekeres A, Tóth G (2003) J Pept Res 62:145–157
35. Grieco P, Giusti L, Carotenuto A, Campiglia P, Calderone V, Lama T, Gomez-Monterrey I, Tartaro G, Mazzoni MR, Novellino E (2005) J Med Chem 48:3153–3163
36. Leitgeb B, Szekeres A (2003) J Mol Struct THEOCHEM 666–667:337–344
37. Gentilucci L, Tolomelli A, Squassabia F (2007) Protein Pept Lett 14:51–56
38. Tömböly C, Kövér KE, Péter A, Tourwé D, Biyashev D, Benyhe S, Borsodi A, Al-Khrasani M, Rónai AZ, Tóth G (2004) J Med Chem 47:735–743