Theor Chem Acc (2010) 125:253–268
DOI 10.1007/s00214-009-0645-6

REGULAR ARTICLE

Hydrogen bonding in the urea dimers and adenine–thymine DNA base pair: anharmonic effects in the intermolecular H-bond and intramolecular H-stretching vibrations

Attila Bende

Received: 16 December 2008 / Accepted: 18 September 2009 / Published online: 6 October 2009

Abstract The equilibrium structures, binding energies, vibrational harmonic frequencies, and the anharmonic corrections for two different (cyclic and asymmetric) urea dimers and for the adenine–thymine DNA base pair system have been studied using the second-order Møller–Plesset perturbation theory (MP2) method and different density functional theory (DFT) exchange–correlation (XC) functionals (BLYP, B3LYP, PBE, HCTH407, KMLYP, and BH and HLYP) with the D95V, D95V**, and D95V++** basis sets. The widely used a posteriori Boys–Bernardi or counterpoise correction scheme for basis set superposition error (BSSE) has been included in the calculations to take into account the BSSE effects during geometry optimization (on structure), on binding energies and on the different levels of approximation used for calculating the vibrational frequencies. The results obtained with the ab initio MP2 method are compared with those calculated with different DFT XC functionals; and finally the suitability of these DFT XC functionals to describe intermolecular hydrogen bonds as well as harmonic frequencies and the anharmonic corrections is assessed and discussed.

Keywords Urea dimer · Adenine–thymine base pairs · Anharmonic corrections · H-stretching vibration · BSSE effect

Dedicated to Professor Sandor Suhai on the occasion of his 65th birthday and published as part of the Suhai Festschrift Issue.

Electronic supplementary material The online version of this article (doi:10.1007/s00214-009-0645-6) contains supplementary material, which is available to authorized users.

A. Bende (✉)
Molecular and Biomolecular Physics Department,
National Institute for R&D of Isotopic and Molecular Technologies, Donath Street, No. 71-103,
400293 Cluj-Napoca, Romania
e-mail: bende@itim-cj.ro

A. Bende
Molecular Biophysics Department,
German Cancer Research Centre,
Im Neuenheimer Feld 580,
69120 Heidelberg, Germany

1 Introduction

Hydrogen bonding is ubiquitous in nature and governs a wide array of chemical and biological processes ranging from local structure in molecular liquids to the global structure and folding dynamics of proteins [1, 2]. Although the hydrogen bond (HB) is well studied, its low-frequency vibrations, the large-amplitude motions involving stretching and bending along the actual (non-covalent) HB coordinates, have been rarely investigated [3]. Information about these vibrations offers exceptional insight into the potential energy surface of the interaction and, therefore, further enhances our understanding of the HB and its impact on molecular structure and dynamics. The "C=O···H–N" type HB is one of the most frequently occurring bonds in the biological systems. They can be found as a main component of the DNA base pairs' interaction systems or in protein α-helix and β-sheets. In this sense, not only the static or averaged molecular structures are important, but also molecular dynamics and dynamical fluctuations are also fundamental for many biological functions; but certainly not all. In our previous work, we have investigated the nature of this C=O···H–N hydrogen bond in the case of the formamide–water and formamide–formamide systems [3] giving a detailed description of binding energies, and of inter- and

intramolecular harmonic frequencies including their anharmonic corrections.

The urea dimer presents, to some extent, a good similarity with the formamide–formamide system. The difference between these two systems is the existence of the second cyclic system in the urea dimer. In the case of the urea dimer, one of the HBs is the weak C–N···H–N hydrogen bond. The planarity (or non-planarity) of the urea dimer system has been the focus of a number of studies. Masunov and Dannenberg [4, 5] considered different levels of theory (HF, MP2, and DFT with and without BSSE corrections), and the most stable conformation was found to be the non-planar structure (using the MP2 method with the D95++** basis set). However, they stressed that inclusion of vibrational and thermal corrections in the calculations of the molecular structure might give an effectively planar structure. Similar results were reported in the case of urea dimer structure also by Belosludov et al. [6] and Skurski et al. [7]. When considering the previous remarks on the geometry of urea dimer, the crystal, and cluster structures of urea were the subject of many theoretical works [5, 8, 9], using ab initio, molecular dynamics or crystal theory methods. In all of these papers, a planar monomer geometry/structure for urea was assumed and the theoretical calculations were undertaken taking this fact into account. This planar monomer geometry/structure was also confirmed by Shukla et al. [10] measuring the Compton profile anisotropies in crystalline urea. We consider the papers of Rousseau and Keuleers [11, 12] as very important works for the use in elucidating the structure of urea (and its dimer) because the detailed descriptions of the vibrational spectra of urea (and its dimer) in both the gas and crystal phases are presented. They concluded that the vibrational analysis of solid urea and of the gas phase of urea are not comparable, which is mostly due to the different planar or non-planar conformation of urea in different states. Motivated by this unusual behavior as well as by the source of the great flexibility of urea, many authors present different molecular properties such as polarizability and hyperpolarizability [13], the excess electron binding to a single [14] or dimer molecular system [15], as a solvent around peptide conformations [16], and hereby giving important contributions on the understanding of urea behavior in different environments.

The A–T base pair's intermolecular interaction can be considered as a contribution of two very important HBs: C=O···H–N and $\gtrsim$N···H–N bonds. Owing to the biological manifoldness of DNA base pair conformations (dry-DNA, wet-DNA, double helix, super-double helix, etc.), their theoretical characterization is much diversified. A given theoretical model should take into account the backbone and environment effects as well as the influence of the DNA–protein interactions. In view of this fact, choosing the correct theoretical method is essential in the investigation of DNA base pairs. In the last 2–3 years, substantial advances in computational modeling have allowed for the inclusion of accurate electron correlation effects in their calculations. Šponer et al. [17] and Podolyan et al. [18] used the well-known MP2 method with different basis sets (mostly including the polarization effects) and they gave a very detailed description of the interaction energies and geometry structures of the adenine–thymine and guanine–cytosine base pairs. They also show that the electron correlation effects could give important contributions to the interaction energy and to the dimer geometry. Owing to the quick growth of the two-electron integrals, it is obvious that the MP2 method could not be applied for the extended structures such as sugar–phosphate chain or the water and metal–ion environments. Similarly, good results can be obtained with much less computational effort using DFT methods. In the work of Guerra et al. [19], not only the simple base pair structure, but also the backbone effects were presented. They compared the efficiency of several DFT functionals to obtain theoretical values close to the experimental results. Considering the experimental data for interaction energies and geometry structures, they conclude that the DFT functionals (especially the BP86) are adequate to describe biologically relevant molecules involving HBs; in particular, the DNA base pair interactions. In addition, a revised version of PM3 semiempirical method presented by Giese et al. [20] could also be considered as another solution for describing large systems. At the same time, the normal mode analysis of molecular vibration could give us supplementary information about the efficiency of different methods. In connection with this, several vibrational frequency calculations and experimental measurements were performed [18, 21–24] to study different interaction effects in normal mode vibrations of base pairs (mostly for guanine–cytosine dimer). For instance, the intermolecular interactions could significantly influence the intramolecular normal mode vibrations, such as red shift or improper blue shifting [25, 26] of the vibrational frequencies. Furthermore, the intermolecular normal mode vibrations depend very much on the applied method or basis sets, and last but not the least on the BSSE effects [3, 27–29].

Anharmonic effects represent significant corrections in the molecular normal mode vibrations involving hydrogen atom [30–32]. Watanabe et al. [32] showed that including anharmonic correction [scaled hypersphere search method (SHS)], all fundamental frequencies in gas phase experiments of intramolecular modes of $(H_2O)_n$ (n = 2–5) were excellently reproduced. At the same time, the vibrational shifts derived from the anharmonic calculations show a good agreement with the trends found in the experiments [33].

The goal of this study is to give an accurate description of intermolecular normal modes and to present different

intermolecular interaction effects that could influence the monomer-type vibrations, considering two urea dimer cases and the adenine–thymine molecular system. Accordingly, several DFT functionals were tested by comparing them with the corresponding MP2 results.

2 Methods and computational details

Regarding the theoretical methods, it is well established that for an accurate description of HBs, ab initio techniques are needed which include the electronic correlation level, i.e. with an error <0.04 eV (1 kcal/mol) to predict the HB strength. Thus, the observed underestimation of the HB strength by the Hartree–Fock (HF) calculations (where the electron correlation effects are missing) is overcome using correlated methods such as second-order Møller–Plesset perturbation theory (MP2) or coupled cluster (CC) methods. Yet, it is necessary to use very large high-quality basis sets to expand the wave function and to get reliable HB properties. This fact and the necessity of correlated methods for accurate description of HBs make such studies computationally too expensive and only applicable to molecular complexes of at most a few tens of atoms. Therefore, strategies to study HBs with similar accuracy as the MP2 or higher levels of theory, but computationally less expensive are needed. In this sense, density functional theory (DFT) is a method that includes electronic correlation. Because of the way DFT is formulated, it emphasizes corrections to an independent particle model. This is usually called dynamic correlations (electron correlation due to the Coulomb hole). In contrast, wavefunction methods (like perturbation, coupled-cluster or multi reference configuration interaction methods) most easily correct for dispersion interaction and for longer range static correlations made when electrons delocalize over a longer distance, effects that DFT cannot treat adequately [34–36]. The dynamic correlation is important in short and medium ranges of interaction, while the static (or non-dynamic, near degeneracy, or sometimes left–right) correlation and the dispersion effects dominate the long-range part. Dispersion interactions in the van der Waals (vdW) complexes depend on fluctuations of the density in the tails of the monomers forming the complex; namely, they depend strongly on a correct description of electron correlation in the vdW region [34]. The lack of non-local static correlation and the long-range dynamic correlation effects (dispersion), respectively, in the LDA and GGA functionals is believed to be responsible for the deficiency of DFT to obtain the long-range vdW interactions [37]. Better results are obtained for vdW complexes dominated by electrostatic forces, such as classical hydrogen bond systems. In this way, the accuracy of DFT to describe the HB interaction relies on the applied functional to approximate the electronic exchange–correlation (XC) contribution. To overcome this problem, first we try to find those XC functionals that could describe HBs with an approximate accuracy as the MP2. In the last few years, a number of significant works could be found in the literature, which compared the efficiency of different DFT functionals with those obtained using the MP2 method. According to this fact, for the first step, we select six XC functionals (BLYP, B3LYP, PBE, HCTH407, KMLYP, and BHandHLYP) [19, 38–40], which could give a satisfactory solution for our problem based on their results presented till now. Hereafter, the BHandHLYP XC functional name will be abbreviated as BHLYP. The HCTH407 represents the Handy's functional including gradient-corrected correlation [41–43], PBE is the gradient-corrected correlation functional of Perdew et al. [44, 45], while KMLYP [40] is a hybrid functional of Kang and Musgrave, which implements the following scheme:

$$\begin{aligned}\text{KMLYP} : {} & (1-a) * E_X^{\text{Slater}} \\ & + a * E_X^{\text{HF}} + (1-b) * E_C^{\text{VWN}} + b * E_C^{\text{LYP}} \end{aligned} \quad (1)$$

where $a = 0.557$ and $b = 0.448$. The Gaussian IOP keywords for the KMLYP functional are:

$$\begin{aligned}&\#\text{iop}(3/76 = 1000005570);\ \text{iop}(3/77 = 0000004430);\\ &\text{iop}(3/78 = 044801000)\end{aligned} \quad (2)$$

Many papers have indicated that the impact of BSSE on the geometries of weakly bound systems is smaller for the DFT methods than that for ab initio methods such as the MP2, but their influence even in case of DFT could not be neglected [28]. Accordingly, we use BSSE elimination technique for all applied methods. The BSSE was corrected using the counterpoise (CP) method [46] as implemented in the Gaussian03 package suite [47]. The uncorrected and BSSE-corrected energies, geometries, harmonic frequencies, and their anharmonic corrections [48, 49] were calculated for the MP2 and DFT levels of theory using the D95V, D95V**, and D95V**++ basis sets [50]. Apart from the fact that these XC functionals can offer us reasonable description compared with the MP2 results, their selection is basically connected with the availability of these functionals included in the C.02 subversion of the Gaussian program package to use them for anharmonic correction calculations. For the harmonic frequencies, no scaling factor was applied. Considering simultaneously, the BSSE and anharmonic corrections on the normal mode vibration analysis, one should make some compromises. In case of anharmonic corrections, one has to perform two different energy calculations including first- and second-order analytical energy derivatives for each normal mode. Furthermore, for each energy job, one should consider five

different single point energy calculations using different basis sets (with and without ghost orbitals it is required in case of counterpoise correction) to consider the BSSE effects. This means that one would have a number of ten energy calculations (considering also the first and second analytical derivatives), for a single normal mode. In this way, when compared with the harmonic frequency calculations, the computer time for the BSSE-free anharmonic corrections would increase very much. It was shown that well-chosen DFT functionals (e.g. B3LYP) [51, 52] coupled with double-zeta quality basis sets [53, 54], including polarization and diffuse functions, represent a good approach for small and medium size molecular systems to achieve the best compromise between computer time and quality of results. Furthermore, based on our previous experience [3], we consider that the D95V basis set has better performance (showing less BSSE errors) than the 6-31G Pople's basis set family.

Nevertheless, to estimate the magnitude of errors given by the medium-quality basis sets and theoretical methods, include higher level correlation methods (e.g. CCSDT) and successively larger shells of polarization functions in the basis set. Considering the local (L) approximation [55–58] and the density-fitting (DF) techniques [59], the computational cost for the classical MP2 method can be drastically reduced. In this way, highly polarized basis sets can be easily included in our calculation without having a significant computational effort. These combined local and density-fitting approximations could also be considered in the case of coupled-cluster theory. All of these approximation techniques are implemented in the Molpro program package suite [60]. For the density fitting, local correlation methods [DF-LMP2 and DF-LCCSD(T)] the basis set superposition errors (BSSE) is drastically reduced [61, 62] and in this way the LMP2 method is suitable for the calculation of intermolecular interactions. Two of the most important advantages of the method are that the potentially expensive counterpoise (CP) corrections are avoided and the dispersion effects, which are crucial for stacking interaction of the aromatic rings, are also included. A very good comparative study in case of DNA base stacking interaction is presented in [63], where it was shown that the LMP2 and the DF-LMP2 intermolecular interaction energies are very close to the standard MP2 values. Considering the local character of occupied and virtual orbitals in the local correlation treatment [64], one can easily obtain also the dispersion part (an intermolecular effect) of the correlation contribution. Another possibility to perform high-quality quantum chemical calculation is to use DFT functionals improved by empirical long-range dispersion corrections [65] or to apply the new generation hybrid meta-GGA (generalized gradient approximation) M06-2x functional developed in the Truhlar group [66]. These methods were recently implemented in the NWChem program package [67]. Considering among others also many DFT functionals presented in this study, a very suggestive test benchmark was presented by Korth and Grimme [68]. They found that the mean absolute deviation obtained with different DFT functionals decreases very strongly from the local density approximation (LDA) to GGAs, but then was less pronounced to meta-hybrid-GGAs (like M06-2x).

The results obtained with the above-mentioned quantum chemistry packages were processed using the Molden [69] and the Gabedit [70] molecular graphics softwares.

3 Results and discussion

In this section, we present the results for interaction energies and geometry parameters of HBs in cyclic and asymmetric urea dimers, as well as for those obtained for the adenine–thymine system calculated at the MP2 and DFT level of theory, using the D95V and D95V**++ basis sets. After the analysis of the DFT and MP2 results, where the MP2 interaction energy and HB distances for urea dimers are compared with all XC functional (BLYP, B3LYP, PBE, HCTH407, KMLYP, BHLYP) results, we focus on the harmonic vibrational frequencies and their anharmonic frequency correction calculations for different normal modes which are affected by the presence of an adjacent molecule. The aim of the MP2 and DFT comparative studies is to find the best correspondence between the MP2 and DFT values to use the adequate XC functionals for further calculations of the adenine–thymine system.

To estimate the differences in intermolecular interaction energy or intermolecular distance which come from applying better correlation methods and larger basis sets, different potential energy curves were drawn. First, using HF, DF-LMP2, DF-LCCSD(T), Truhlar's M06-2x DFT functional, PBE-D, and BLYP-D methods with cc-pVTZ basis set, while in the second case applying the DF-LMP2 method with cc-pVXZ and Aug-cc-pVXZ (where X = D, T, Q) basis sets [71, 72]. Their effects (high correlation contribution and basis set dependence) for different urea dimer structures will be also discussed.

Furthermore, in the Electronic Supplementary Material (ESM), a large amount of data is presented including harmonic and anharmonic frequencies, diagonal and off-diagonal anharmonic constants in table format, as well as different normal mode vibrations in graphics format, for both of urea dimer and the adenine–thymine base pair.

3.1 The urea dimer

3.1.1 Geometry structure of urea dimers

In Table 1, we list the interaction energies (kcal/mol) and intermolecular HB distances (Å) obtained for cyclic and asymmetric urea dimer structures with the MP2 method and five different DFT XC functionals (both in uncorrected and CP-corrected cases) using the D95V**++ basis set. In the last five rows of Table 1, the same interaction energy and intermolecular HB distance obtained with DF-LMP2 method using the D95V**++ and cc-pVQZ basis sets are also shown as well as the interaction energy obtained using DF-LCCSD(T) method. The cyclic urea dimer is bonded by two equivalent C=O···H–N HBs presented in Fig. 1, while the asymmetric one is formed by two different C=O···H–N and $\gtrsim$N···H–N bonds as shown in Fig. 2. In both cyclic and asymmetric dimer cases, the urea monomers present a non-planar geometry configuration.

Table 1 The ε interaction energies (kcal/mol) and R intermolecular distances (Å) in the case of cyclic and asymmetric urea dimer structures, optimized at different levels of theory and considering the D95V**++ and cc-pVQZ basis sets

Method	Cyclic		Asymmetric		
	ε	$R_{O\cdots H}$	ε	$R_{O\cdots H}$	$R_{N\cdots H}$
MP2					
NoCP[a]	−16.53	1.86	−14.44	2.10	2.178
CP[b]	−13.18	1.93	−11.43	2.19	2.30
BLYP					
NoCP	−14.15	1.84	−10.54	1.95	2.11
CP	−13.52	1.86	−9.77	1.99	2.16
B3LYP					
NoCP	−15.04	1.84	−11.35	1.96	2.13
CP	−14.43	1.85	−10.64	2.01	2.17
PBE					
NoCP	−16.81	1.79	−13.10	1.90	2.03
CP	−16.12	1.80	−12.20	1.91	2.07
HTCH407					
NoCP	−12.33	1.96	−9.48	2.12	2.33
CP	−11.76	1.98	−8.92	2.14	2.39
KMLYP					
NoCP	−20.74	1.75	−16.81	1.93	2.06
CP	−20.08	1.75	−16.06	1.94	2.10
BHLYP[c]					
NoCP	−15.59	1.81	−12.15	2.01	2.18
CP	−15.01	1.86	−11.52	2.03	2.22
DF-LMP2	−13.26	1.93	−10.00	2.07	2.22
E_{disp}^{LMP2}	−2.21	–	−2.10	–	–
DF-LMP2/vqz[d]	−14.83	1.86	−12.39	2.00	2.15
E_{disp}^{LMP2}/vqz	−3.56	–	−3.81	–	–
DF-LCCSD(T)[e]/vqz	−14.29	–	−11.82	–	–

[a] Without counterpoise correction

[b] With counterpoise correction

[c] BHandHLYP

[d] cc-pVQZ

[e] The LMP2 optimized geometry was used

The interaction energy and HB distance values show that a good fitting of DFT values with the MP2 results is not obvious for any of the selected XC functionals. In the case of cyclic structure, one can observe that the best agreement for the intermolecular interaction energies and intermolecular $R_{O\cdots H}$ distances is given by the BHLYP and B3LYP XC functionals. The interaction energy values are −15.035 kcal/mol for B3LYP, −15.586 kcal/mol for BHLYP and −16.534 kcal/mol for MP2, while the intermolecular distance for $R_{O\cdots H}$ is 1.8377 (B3LYP), 1.8514 (BHLYP), and 1.8562 Å (MP2). At the same time, reasonable values could be also obtained applying the BLYP functional, while in the case of KMLYP, PBE, and HTCH407 the energy and geometry parameter results are quite different from the MP2 values. Similar conclusions can be drawn for the asymmetric structure of urea, but in this case the accordance between the DFT and MP2 results is not so obvious even between B3LYP, BHLYP XC functionals, and the MP2 values. Regarding the "counterpoise" corrected interaction energy and intermolecular distance results; one can see that the BSSE corrections for both cyclic and asymmetric structures are more relevant at the MP2 level, than those for the DFT functionals.

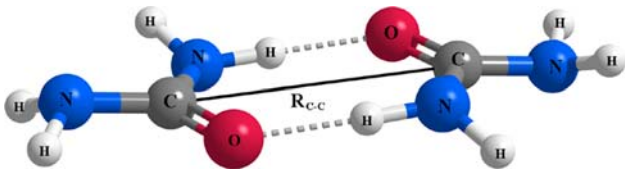

Fig. 1 The equilibrium geometry structure of cyclic urea dimer obtained at local-MP2 level of theory, using the cc-pVTZ basis set

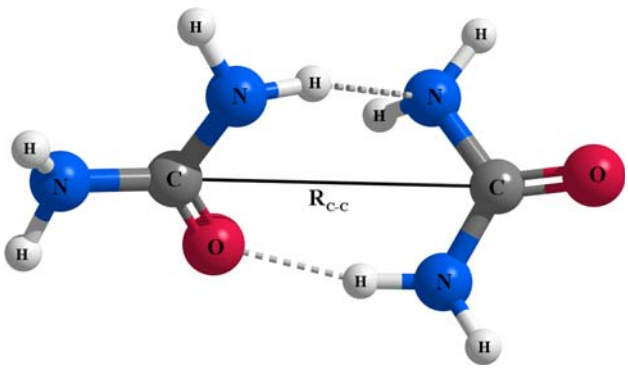

Fig. 2 The equilibrium geometry structure of asymmetric urea dimer obtained at local-MP2 level of theory, using the cc-pVTZ basis set

In Table 2, the same interaction energies (kcal/mol) and intermolecular HB distances (Å) are collected as presented before, but obtained only with the MP2 method, and B3LYP, and BHLYP DFT XC functionals (both in uncorrected and CP-corrected cases) and considering the D95V, D95V** and D95V**++ basis sets. Analyzing the geometry structure and calculating the intermolecular interaction energies for both cyclic and asymmetric urea dimers using the D95V basis set, but considering different polarization and diffuse functions, some basic remarks can be made. First, there is a basic difference between the geometry structure obtained with the D95V basis set and those obtained with the D95V** and D95V**++ basis sets. In the first case, we got a totally planar configuration of both urea monomers from the dimer structure, while in the second case, they show non-planar configurations. The intermolecular interaction energies obtained with the simple D95 basis set show larger minima than those for the D95V** and D95V**++ basis sets, while the difference between the D95V** and D95V**++ values is not so significant.

Table 2 The ε interaction energies (kcal/mol) and R intermolecular distances (Å) in the case of cyclic and asymmetric urea dimer structures, optimized at different levels of theory and considering the D95V, D95V** and D95V**++ basis sets

Method	Cyclic		Asymmetric		
	ε	$R_{O\cdots H}$	ε	$R_{O\cdots H}$	$R_{N\cdots H}$
D95V					
MP2					
NoCP[a]	−19.58	1.84	−16.01	1.98	2.15
CP[b]	−14.71	1.94	−11.93	2.14	2.36
B3LYP					
NoCP	−21.73	1.74	−12.97	1.98	2.18
CP	−18.91	1.78	−11.14	2.07	2.25
BHLYP					
NoCP	−22.50	1.75	−16.07	2.00	2.26
CP	−19.95	1.79	−14.93	2.06	2.33
D95V**					
MP2					
NoCP	−16.06	1.82	−14.07	1.96	2.08
CP	−12.32	1.95	−10.58	2.05	2.21
B3LYP					
NoCP	−16.83	1.82	−15.46	1.93	2.09
CP	−14.80	1.85	−13.36	1.94	2.13
BHLYP					
NoCP	−16.92	1.84	−13.38	1.98	2.15
CP	−15.32	1.86	−11.93	2.00	2.19
D95V++**					
MP2					
NoCP	−16.53	1.86	−14.44	2.10	2.18
CP	−13.18	1.93	−11.43	2.19	2.30
B3LYP					
NoCP	−15.04	1.84	−11.35	1.96	2.13
CP	−14.43	1.85	−10.64	2.01	2.17
BHLYP					
NoCP	−15.59	1.81	−12.15	2.01	2.18
CP	−15.01	1.86	−11.52	2.03	2.22

[a] Without counterpoise correction

[b] With counterpoise correction

When comparing the amounts of the total interaction energy and its dispersion part obtained by the DF-LMP2 method with two different basis sets (D95V**++ and cc-pVQZ), it was found that the dispersion contribution gives around 16–24% from the total energy value. This is −2.21 kcal/mol for the cyclic urea structure and −2.10 kcal/mol for the asymmetric urea, if one applies the D95V**++ basis set, and −3.56 and −3.81 kcal/mol, respectively, for the cc-pVQZ basis set.

When considering the classical MP2 results obtained with the D95V**++ basis set (see Table 1), one could observe that for the cyclic geometry, the ε interaction energy and R intermolecular HB distance are quite close to the corresponding value calculated with the DF-LMP2 method, while for asymmetric structure these two methods show discrepancies. If we compare the same ε and R values for both cyclic and asymmetric structures at the DF-LMP2 level of theory, but obtained with different basis sets (D95V**++ vs. cc-pVQZ), one could see that the energy results increase by 1.58 and 2.39 kcal/mol, respectively, while for the R distance we get smaller values with 0.06 and 0.07 Å, respectively. Focusing on the dispersion part of the intermolecular interaction energy, significant dispersion energy growth can be found (1.35 kcal/mol for cyclic dimer and 1.71 kcal/mol for asymmetric case), when the larger cc-pVQZ basis set is used against the D95V**++. These represent the major contribution to the total interaction energy increase. To consider higher correlation effects, other than those included in the MP2, the interaction energies were computed applying the DF-LCCSD(T) method, using the same cc-pVQZ basis set and taking into account the DF-LMP2 optimized geometry. Higher correlation effects give positive energy contribution. These are 0.54 kcal/mol for cyclic structure and 0.57 kcal/mol in asymmetric case, but compared with the whole magnitude of the energy value, it is not as significant as the dispersion effect.

The HB is primarily electrostatic in nature, where the proton's electron cloud is disclosed by the donor atom and the proton's positive charge is attracted electrostatically by the so-called lone pair (non-bonded) electrons of the acceptor atom. This attractive force is complemented by quantum chemical contributions including exchange interaction that is repulsive, and polarization and charge-

transfer components, which are attractive. At the electron correlation level, this forces are completed with other contributions such as dispersion interaction. DFT includes XC via XC functionals. For the hybrid functionals, such as B3LYP or BHLYP, exchange effects are turned down by the admixture of both Slater and HF type exchange, however, not to reproduce exact exchange energies, but to provide better total correlation energy [34]. It was shown by Cremer [34] that the hybrid XC functionals, constructed in this manner, slightly underestimate the exchange part and a shade better overestimate the correlation part. In this way, the difference could partially cover other energy contributions (i.e., dispersion effects) that are not included in this XC functionals by construction. Accordingly, one can consider that the dispersion effects, which are missing from the B3LYP and BHLYP functionals, are partially compensated by this energy surplus. This could explain the facts that why the B3LYP and BHLYP results are quite close to the MP2 or LMP2 values. In case of the stacking interaction, where the correlation covered by the hybrid XC functionals is small, the energy surplus is not able to compensate anymore the dispersion effects.

In Figs. 3 and 4, several potential energy curves are shown for cyclic and asymmetric urea dimers, respectively. They were obtained with HF, density-fitting local correlation methods [LMP2 and LCCSD(T)], Truhlar's M06-2x

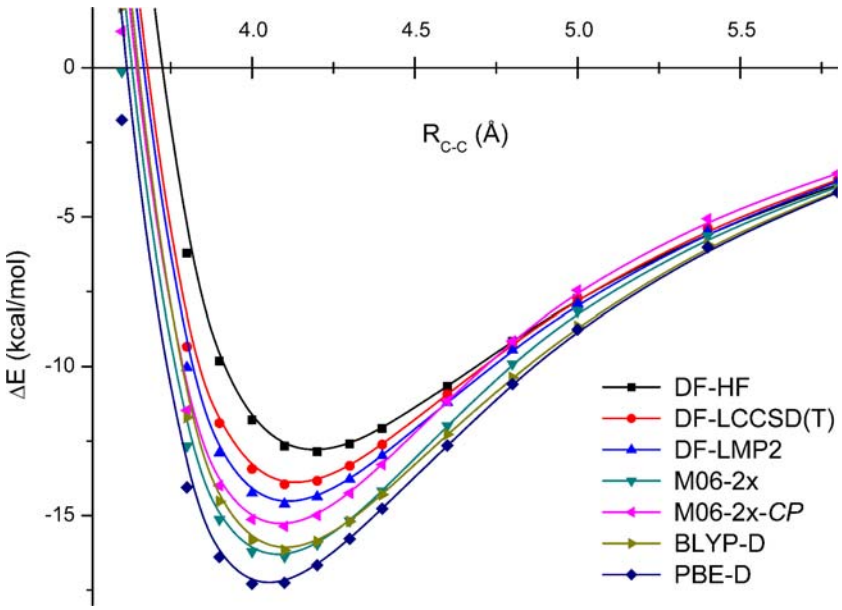

Fig. 3 The intermolecular potential energy curves (kcal/mol) of cyclic urea dimer obtained at HF, LMP2, LCCSD(T), M06-2x, counterpoise corrected M06-2x, BLYP-D and PBE-D (D = empirically corrected XC functionals by dispersion effects) levels of theory using the cc-pVTZ basis set

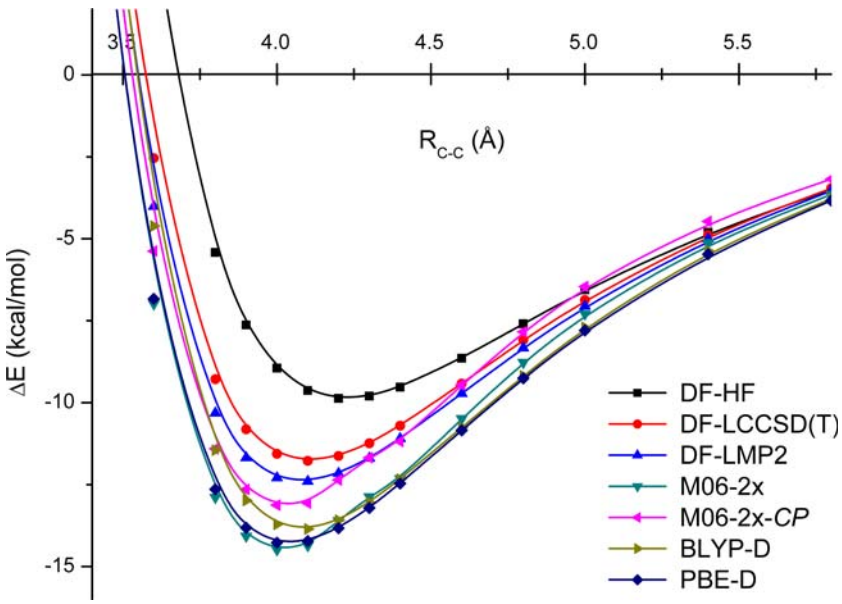

Fig. 4 The intermolecular potential energy curves (kcal/mol) of asymmetric urea dimer obtained at HF, LMP2, LCCSD(T), M06-2x, counterpoise corrected M06-2x, BLYP-D, and PBE-D (D = empirically corrected XC functionals by dispersion effects) levels of theory using the cc-pVTZ basis set

hybrid meta-GGA DFT functional, as well as at BLYP and PBE XC functionals with empirical dispersion correction methods and with the cc-pVDZ basis set. On inspecting the potential energy profiles, it can be seen that the DFT functionals gave lower minima (about 1–1.5 kcal/mol) compared with local correlation methods [DF-LMP2 and DF-LCCSD(T)]. This discrepancy could be explained, partially, with the lack of BSSE correction in the case of above enumerated DFT functionals, as it also could be observed for those presented in the first part of this subsection (see also Table 1). Accordingly, for the M06-2x functional, the BSSE correction increases the minimum of the potential energy curve with 1.02 kcal/mol for the cyclic case and with 1.36 kcal/mol for asymmetric dimer (see the M06-2x and M06-2x-CP curves in Figs. 2 and 3).

To draw inferences about the basis set dependency, different potential energy curves (Figs. 5, 6) were taken. Here, we have considered the DF-LMP2 method used together with cc-pVXZ and Aug-cc-pVXZ (where $X = D$, T, Q) basis sets. The potential energy profiles show that the double-zeta quality basis sets (cc-pVDZ and Aug-cc-pVDZ) gave partly different energy curves from the results obtained with triple- or quadruple-zeta quality basis sets, but the deviation is <1.5 kcal/mol.

Summarizing the contributions given by different electron correlation methods and basis sets, it can be stated that

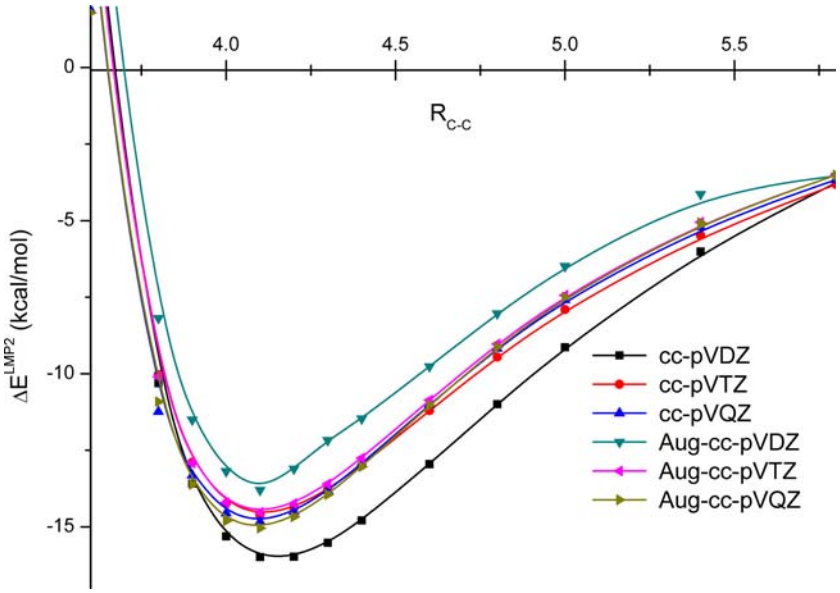

Fig. 5 The intermolecular potential energy curves (kcal/mol) of cyclic urea dimer obtained with LMP2 method and considering the cc-pVXZ and Aug-cc-pVXZ (where X = D, T, Q) basis sets

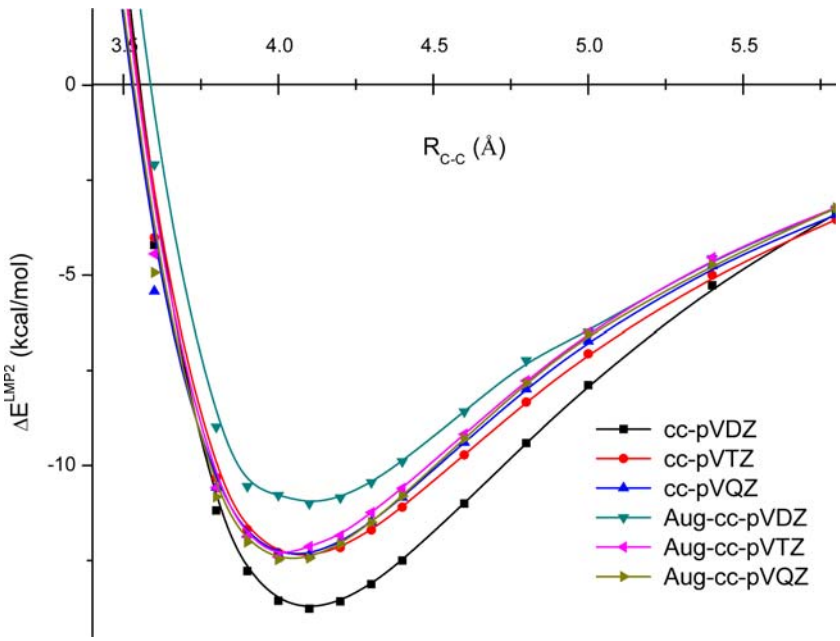

Fig. 6 The intermolecular potential energy curves (kcal/mol) of asymmetric urea dimer obtained with LMP2 method and considering the cc-pVXZ and Aug-cc-pVXZ (where X = D, T, Q) basis sets

the most important correction for interaction energies and intermolecular distances is obtained taking into account the dispersion effect. This is followed by the basis set quality induced corrections and finally by the third- and fourth-order electron correlation contributions.

3.1.2 Normal mode vibrations of urea dimers

Each urea dimer has 42 vibrational normal modes (see Table 3) from which a number of 36 (18 for each monomer) vibrations are characteristic to monomer-type vibrational motion. The last six normal modes with the lowest frequency values have purely intermolecular character. These 18 monomer frequencies of dimer structure are split in to a form of frequency pairs (doublets). As compared to the individual monomer lines they present different frequency shifts and their magnitudes depend very much on the dimer molecular symmetry (cyclic or asymmetric). In ESM Tables 1 and 2, the harmonic frequencies (doublets) and their anharmonic corrections for those monomer-type normal modes of which vibrational motions are substantially perturbed by the adjoining molecule were collected, both for the cyclic and asymmetric urea dimer structures. They were determined in advance by identifying the frequency value that corresponds to each normal mode vibration and by visualization of their vibrational characters [69]. This preliminary vibrational analysis shows that the most shifted frequencies of the monomer-type vibrations are the covalent N–H bond stretching and H–N out-of-plane vibrations (the H atom is involved in the formation of HB) as well as the C=O double bond stretching vibrations. In ESM Tables 1 and 2 the dimer-type normal mode frequencies are also compared with those obtained for the individual monomer vibrations, where in addition the intermolecular BSSE corrections for dimer structures were considered. It was found that, the ν_1, ν_2, and ν_3 normal modes are N–H bond stretching vibrations where the H atoms do not take part in intermolecular HB formations. The ν_4, ν_5, and ν_6 are one N–H and two C=O bond stretching vibrations located at the N–H···O=C intermolecular HBs (see Figs. 1 and 2). In the case of geometry structure and intermolecular interaction energy, it was obtained that the best agreement between the MP2 and DFT results is given by B3LYP and BHLYP XC functionals. Therefore, in ESM Tables 1 and 2 the harmonic frequency values and anharmonic frequency corrections are presented only for these methods. The efficiency of B3LYP XC functional in describing urea dimer structures was also pointed out by the authors of reference [6].

It can be observed that in the case of cyclic urea structure, the doublet frequency splits into ν_1, ν_2, and ν_3 normal modes that are almost irrelevant. More precisely, the frequency shifts, which are induced by each monomer on the adjoining molecules, have the same magnitude. Only those normal modes present different frequency shifts and implicitly larger doublet splits where the perturbation of HB vibration is present (ν_4, ν_5, and ν_6). As can be seen in ESM Table 1, the dimer frequency shift could be attributed to several effects like anharmonic corrections, BSSE effects, or intermolecular effects. It should be mentioned that BSSE is not a real physical effect and normally it must be considered together with the intermolecular effects. However to see how important frequency shifts could induce the BSSE error, we consider as a separate effects. Accordingly, in the following part, we try to describe each of these effects separately and in more detail. First, let us discuss the role of anharmonic effects. In the case of N–H bond stretching vibrations, the magnitudes of anharmonic corrections (frequency red shift) are quite large ($\approx$150–170 cm^{-1} for ν_1, ν_2, and ν_3, and >200 cm^{-1} for ν_4, respectively) while in the case of a_1, a_2, and a_3 anharmonic corrections, for both of doublet frequency values, the magnitude of frequency shifts are the same. For a_4 correction, the size of doublet frequency shifts is different (e.g., at MP2 level $\nu_4' - a_4' = 3,457.8 - 3,229.3\,\mathrm{cm}^{-1} = 228.5\,\mathrm{cm}^{-1}$ and $\nu_4'' - a_4'' = 3,419.6 - 3,149.0\,\mathrm{cm}^{-1} = 270.6\,\mathrm{cm}^{-1}$). Regarding ν_5 and ν_6, C=O stretching modes one can see that the anharmonic corrections are much smaller than in the H–N case, their shifts are about 30–40 cm^{-1} and the behavior of their frequency split is similar to ν_4 mode. The BSSE corrections could also be considered as another important effect. Having intermolecular character, it can be explicitly observed that considerable BSSE corrections are obtained only for ν_4, ν_5, and ν_6 normal modes. Both for harmonic and anharmonic approximations only the MP2 results show large frequency corrections. Moreover, the frequency of H–N stretching mode is more affected by the BSSE than the C=O stretching vibrations. The intermolecular effects (or molecular association effects) present another source of frequency shifts. Here, one should compare the monomer-type frequency (harmonic or anharmonic) values from the dimer structure with the corresponding individual monomer frequencies. The results show that those normal modes, where the vibration motion is far from the intermolecular interaction region (ν_1, ν_2, and ν_3) are less perturbed than the ν_4, ν_5, and ν_6 normal modes where the H–N and C=O bonds are the integral part of the HBs. In case of ν_4, we have for MP2 181.1 and 219.3 cm^{-1}, for B3LYP 242.5 and 285.3 cm^{-1} and for BHLYP 204.3 and 240.3 cm^{-1} frequency red shifts. Similar effects can be found for ν_5 and ν_6 C=O stretching modes, but this being a strong double covalent bond the frequency shifts are smaller. The correction scheme for ν_4 and ν_5 is as follows (see also ESM Fig. 1):

Table 3 The harmonic (v) frequencies (cm^{-1}) of the normal mode vibrations of cyclic and asymmetric urea dimers obtained at the MP2, LMP2 and B3LYP levels of theory (with BSSE correction for MP2 and B3LYP methods), using the D95V++** and cc-pVQZ basis sets

Nr.	Cyclic				Asymmetric			
	MP2	LMP2	B3LYP	LMP2 (cc-pVQZ)	MP2	LMP2	B3LYP	LMP2 (cc-pVQZ)
v_1								
1	3,784.1	3,783.4	3,712.6	3,738.4	3,795.0	3,780.5	3,718.2	3,745.5
2	3,784.0	3,783.2	3,712.6	3,738.3	3,777.2	3,766.0	3,697.7	3,733.8
v_2								
3	3,759.4	3,773.2	3,675.9	3,712.6	3,770.7	3,749.0	3,681.3	3,715.4
4	3,758.6	3,772.6	3,675.3	3,712.1	3,737.5	3,734.1	3,664.5	3,694.6
v_3								
5	3,651.1	3,649.0	3,590.2	3,613.2	3,659.3	3,646.6	3,594.4	3,617.7
6	3,651.0	3,649.0	3,590.1	3,613.1	3,622.5	3,605.0	3,544.5	3,567.0
v_4								
7	3,509.0	3,514.6	3,348.9	3,412.4	3,603.3	3,585.8	3,479.0	3,533.9
8	3,480.0	3,484.5	3,307.3	3,373.4	3,584.9	3,562.0	3,461.3	3,505.2
v_5								
9	1,801.2	1,794.7	1,769.4	1,792.0	1,812.7	1,810.6	1,794.4	1,808.7
10	1,781.0	1,774.8	1,744.3	1,765.0	1,788.5	1,789.3	1,756.1	1,774.3
v_6								
11	1,693.1	1,685.2	1,659.1	1,669.2	1,679.9	1,682.2	1,644.6	1,654.2
12	1,687.4	1,680.6	1,653.5	1,666.9	1,675.1	1,673.2	1,638.4	1,644.1
v_7								
13	1,671.0	1,669.8	1,629.4	1,640.3	1,666.5	1,662.2	1,628.2	1,638.0
14	1,666.3	1,662.5	1,622.9	1,635.9	1,657.7	1,656.2	1,622.2	1,628.2
v_8								
15	1,484.4	1,481.2	1,466.3	1,478.0	1,473.6	1,464.3	1,450.7	1,466.1
16	1,469.7	1,465.6	1,448.5	1,462.2	1,428.8	1,423.8	1,401.8	1,418.0
v_9								
17	1,216.2	1,204.9	1,186.6	1,197.5	1,234.9	1,221.8	1,200.6	1,207.3
18	1,213.5	1,202.3	1,181.2	1,193.6	1,205.9	1,208.2	1,178.1	1,190.5
v_{10}								
19	1,077.1	1,071.5	1,065.4	1,077.7	1,092.3	1,101.0	1,081.4	1,083.9
20	1,077.0	1,071.4	1,064.8	1,077.6	1,068.3	1,084.6	1,051.3	1,066.2
v_{11}								
21	993.4	989.6	984.6	996.4	989.9	980.2	978.1	991.2
22	989.0	985.0	980.8	991.4	966.3	954.3	943.3	960.4
v_{12}								
23	801.0	777.0	823.1	808.9	826.1	827.4	813.9	805.7
24	784.1	767.0	781.3	788.2	782.5	781.8	776.2	793.2
v_{13}								
25	746.1	746.7	757.5	759.8	738.1	740.3	735.0	731.5
26	721.7	713.8	756.4	740.5	663.9	662.1	677.5	670.5
v_{14}								
27	596.1	596.4	598.3	602.1	598.2	639.0	634.9	624.1
28	593.7	593.1	588.7	594.7	586.0	598.3	582.3	588.3
v_{15}								
29	572.1	565.8	542.9	551.7	576.7	582.0	571.7	576.3
30	571.6	564.3	538.9	547.3	555.2	574.1	530.5	536.1

Table 3 continued

Nr.	Cyclic				Asymmetric			
	MP2	LMP2	B3LYP	LMP2 (cc-pVQZ)	MP2	LMP2	B3LYP	LMP2 (cc-pVQZ)
ν_{16}								
31	524.5	519.5	515.3	521.0	535.4	546.4	519.7	516.7
32	513.3	503.1	495.9	501.4	507.1	509.3	493.9	492.8
ν_{17}								
33	495.4	437.2	451.1	448.7	493.2	493.2	485.4	475.9
34	489.3	433.0	448.3	446.3	489.7	476.3	447.5	443.5
ν_{18}								
35	308.9	232.2	353.6	367.9	467.4	449.6	435.6	441.7
36	306.4	231.1	351.2	364.1	303.5	380.5	345.2	365.7
ν_{I}								
37	141.5	139.5	155.2	156.0	147.7	139.5	145.1	154.4
ν_{II}								
38	135.2	134.7	147.4	149.3	122.4	125.2	130.3	136.4
ν_{III}								
39	121.9	122.7	135.0	124.7	106.2	102.9	103.0	105.0
ν_{IV}								
40	89.8	58.1	93.8	89.6	74.9	77.9	80.0	76.1
ν_{V}								
41	62.0	45.3	66.9	62.4	66.5	50.4	50.6	55.5
ν_{VI}								
42	38.0	37.2	44.1	37.5	42.1	42.8	39.8	44.0

$$\nu_4' :$$
$$\nu^{\mathrm{mon}} = 3{,}638.9\,\mathrm{cm}^{-1} \xrightarrow[-181.1]{\mathrm{dimer}} \nu^{\mathrm{dim}} \xrightarrow[-228.5]{\mathrm{anh}} \nu^{\mathrm{anh}} \xrightarrow[+195.5]{\mathrm{BSSE}} \nu^{\mathrm{BSSE}} = 3{,}388.8\,\mathrm{cm}^{-1}$$

$$\nu_4'' :$$
$$\nu^{\mathrm{mon}} = 3{,}638.9\,\mathrm{cm}^{-1} \xrightarrow[-219.3]{\mathrm{dimer}} \nu^{\mathrm{dim}} \xrightarrow[-270.6]{\mathrm{anh}} \nu^{\mathrm{anh}} \xrightarrow[+221.4]{\mathrm{BSSE}} \nu^{\mathrm{BSSE}} = 3{,}370.4\,\mathrm{cm}^{-1}$$

$$\nu_5' :$$
$$\nu^{\mathrm{mon}} = 1{,}853.2\,\mathrm{cm}^{-1} \xrightarrow[-54.9]{\mathrm{dimer}} \nu^{\mathrm{dim}} \xrightarrow[-33.4]{\mathrm{anh}} \nu^{\mathrm{anh}} \xrightarrow[+0.7]{\mathrm{BSSE}} \nu^{\mathrm{BSSE}} = 1{,}765.6\,\mathrm{cm}^{-1}$$

$$\nu_5'' :$$
$$\nu^{\mathrm{mon}} = 1{,}853.2\,\mathrm{cm}^{-1} \xrightarrow[-75.3]{\mathrm{dimer}} \nu^{\mathrm{dim}} \xrightarrow[-48.8]{\mathrm{anh}} \nu^{\mathrm{anh}} \xrightarrow[+11.8]{\mathrm{BSSE}} \nu^{\mathrm{BSSE}} = 1{,}740.9\,\mathrm{cm}^{-1}$$

Considering the scheme of frequency corrections by dimer, anharmonic, and BSSE effects presented above, one could observe that they bring different contributions in the case of the ν_4 monomer frequency value (3,638.9 cm^{-1}), where finally we got a double-split frequency pair $\left(\nu_4' \text{ and } \nu_4''\right)$ of 3,370.4 and 3,388.8 cm^{-1} values (see ESM Fig. 1). A similar case can be found for the ν_5 monomer frequency value where we have two frequency values with 24.7 cm^{-1} distance between them.

The asymmetric urea structure contains two different HBs, therefore, the doublet frequency split is a collective effect of these two HBs. We chose six normal modes as shown in ESM Table 2, representing four H–N and two C=O stretching normal modes. They could be classified in the following way: ν_1 and ν_3 are H–N stretching vibrations and are located outside the intermolecular interaction region, ν_2 and ν_4 are also H–N stretching normal modes but inside the intermolecular region, while ν_5 and ν_6 are C=O stretching modes and are found in the same intermolecular interaction zone. Anharmonic effects gave the most important corrections, as it was also concluded above for the cyclic case. The inner HB vibrations located in the intermolecular interaction region can also be easily identified by the BSSE corrections. Usually, they have larger BSSE corrections than those frequency values of which normal mode vibrations are not present in the intermolecular interaction region (ν_2': 9.2 cm^{-1}, ν_2'': 8.9 cm^{-1}, ν_4': 23.4 cm^{-1}, ν_4'': 58.1 cm^{-1} at MP2 level). Furthermore, due to the symmetry breaking in the asymmetric dimer structure, ν_3'' shows also a considerable BSSE correction (12.4 cm^{-1}), while its frequency pair (ν_3') has a very small correction. Similar statement is true for ν_5 normal modes, where ν_5'' presents BSSE effects, but for ν_5' this effect is missing. The anharmonic effect has no significant change with respect to the molecular surrounding; the

anharmonic shifts are between 150 and 170 cm^{-1} for $\nu_1 - \nu_4$ and between 35 and 40 cm^{-1} for ν_5 and ν_6. The correction scheme for ν_4 and ν_5 is as follows (see ESM Fig. 2):

$$\nu_4' : \quad \nu^{\mathrm{mon}} = 3{,}638.9\,\mathrm{cm}^{-1} \xrightarrow[-104.0]{\mathrm{dimer}} \nu^{\mathrm{dim}} \xrightarrow[-157.1]{\mathrm{anh}} \nu^{\mathrm{anh}} \xrightarrow[+33.5]{\mathrm{BSSE}} \nu^{\mathrm{BSSE}} = 3{,}456.3\,\mathrm{cm}^{-1}$$

$$\nu_4'' : \quad \nu^{\mathrm{mon}} = 3{,}638.9\,\mathrm{cm}^{-1} \xrightarrow[-157.1]{\mathrm{dimer}} \nu^{\mathrm{dim}} \xrightarrow[-121.4]{\mathrm{anh}} \nu^{\mathrm{anh}} \xrightarrow[+24.3]{\mathrm{BSSE}} \nu^{\mathrm{BSSE}} = 3{,}429.7\,\mathrm{cm}^{-1}$$

$$\nu_5' : \quad \nu^{\mathrm{mon}} = 1{,}823.2\,\mathrm{cm}^{-1} \xrightarrow[-10.6]{\mathrm{dimer}} \nu^{\mathrm{dim}} \xrightarrow[-37.3]{\mathrm{anh}} \nu^{\mathrm{anh}} \xrightarrow[+2.6]{\mathrm{BSSE}} \nu^{\mathrm{BSSE}} = 1{,}777.9\,\mathrm{cm}^{-1}$$

$$\nu_5'' : \quad \nu^{\mathrm{mon}} = 1{,}823.2\,\mathrm{cm}^{-1} \xrightarrow[-40.9]{\mathrm{dimer}} \nu^{\mathrm{dim}} \xrightarrow[-35.5]{\mathrm{anh}} \nu^{\mathrm{anh}} \xrightarrow[+10.7]{\mathrm{BSSE}} \nu^{\mathrm{BSSE}} = 1{,}757.5\,\mathrm{cm}^{-1}$$

Similarly to the previous cyclic case, for both normal modes, the dimer and anharmonic effects induce negative frequency shifts, while BSSE effects gave smaller, but not negligible positively shifted frequency values.

Apart from the monomer-type frequencies, the molecular association induces a group of another six normal mode vibrations which can be called *intermolecular normal modes*. They can be found in the very-far region (10–250 cm^{-1}) of the molecular IR spectra and show the relative vibrations of two "rigid" urea monomers according to the six degrees of freedom that derive from the intermolecular coordinates. The frequency values of these intermolecular normal mode vibrations are shown in ESM/Table 3 obtained both at the MP2 and DFT (B3LYP and BHLYP) levels of theory. They have strictly intermolecular character and show only anharmonic and BSSE corrections. Scrutinizing the results of normal mode vibrations for the cyclic dimer one can observe that there are three frequencies (see ESM/Fig. 5) of which motion take place in the supermolecular plane (ν_{I}–ν_{III})—let us call them "in-plane" vibrations, while the another three intermolecular normal modes (ν_{IV}–ν_{VI}) show "out-of-plane" vibrations. If we consider together the amount of anharmonic and BSSE corrections, one can see that the "in-plane" vibrations are more affected by these errors than "out-of-plane" normal modes. Yet analyzing separately for only one mode, it can be seen that they have almost similar magnitude, becoming equally relevant corrections for the intermolecular normal mode vibrations. At the same time, if we consider the off-diagonal vibration couplings of these normal modes, one can find that frequencies which belong to the group of "in-plane" or "out-of-plane" their vibrations are strongly coupled inside the group, but much less coupled between the groups. In the case of asymmetric urea dimer the separation of the above-mentioned "in-plane" and "out-of-plane" group of vibrations is not so obvious (see ESM Fig. 3), but similarly to the cyclic dimer, the BSSE and the anharmonic effects are equally relevant. In both cases of dimer structures, the collective effect of BSSE and anharmonic corrections presents about 10–15% from the frequency values.

The *intra–intermolecular normal mode couplings* could give us more detailed information about how strongly the intra- and intermolecular normal modes are coupled. Because the intra- and intermolecular normal modes have very different vibrational frequency values, one should obtain a strong coupling only in some special cases. Analyzing the anharmonic coupling matrix, we found that in case of cyclic dimer the ν_{I}–ν_{III} intermolecular normal modes are coupled only with ν_4' and ν_4'' (ex. $x_(\nu_4'' - \nu_{I)} = +6.73$ cm^{-1}, see ESM Table 4) intramolecular vibrations, while couplings between ν_5' and ν_5'' intranormal modes and ν_{I}–ν_{III} internormal modes are almost missing. This could be explained with the fact that molecular vibrations are performed in the same molecular plane and along the same vibrational direction. In the case of asymmetric urea dimer neither for ν_4' and ν_4'', nor for ν_5' and ν_5'' no significant anharmonic coupling with the intermolecular normal modes could be found (see ESM/Table 5). In this case, the only relevant coupling is $x\left(\nu_5'' - \nu_{II}\right) = -3.86\,\mathrm{cm}^{-1}$.

If one compares the selected theoretical normal mode frequencies with the experimental data (ESM Table 1 and 2), the more appropriate values are obtained for the CP-corrected anharmonic frequencies, both for cyclic and asymmetric structures. At the same time, it can be seen that the dimer approximation is not satisfactory to reproduce these experimental data with the desired high accuracy. Thus, other extended theoretical calculations where also solvent effects and larger cluster sizes are taken into account would be absolutely necessary. In addition, other spectroscopic data, such as vibrational absorption intensities calculated via atomic polar tensor [73] and comparing them with the measured vibrational intensities would be also useful to explain the unusual behavior of urea in gas or solid phase.

3.2 The adenine–thymine dimer

3.2.1 *Normal mode vibrations of the adenine–thymine dimer*

In Table 4 the intermolecular interaction energies and intermolecular HB distances are presented optimized at B3LYP and BHLYP levels of theory (with and without

Table 4 The ε interaction energies (kcal/mol) and R intermolecular distances (Å) in the case of the adenine–thymine DNA base pair, optimized at the MP2, B3LYP and BHLYP levels of theory (with and without counterpoise correction) and considering the D95V and D95V++** basis sets

Method/basis set	ε	$R_{O\cdots H}$	$R_{N\cdots H}$
MP2/D95V	−22.16	1.91	1.72
MP2-CP/D95V	−14.84	2.00	1.84
B3LYP/D95V	−21.88	1.83	1.65
B3LYP-CP/D95V	−18.10	1.87	1.72
B3LYP/D95V++**	−14.96	1.90	1.80
B3LYP-CP/D95V++**	−13.75	1.91	1.82
BHLYP/D95V++**	−14.99	1.93	1.82
BHLYP-CP/D95V++**	−13.92	1.94	1.85

counterpoise correction) and considering the D95V and D95V++** basis sets. Comparing the results obtained using the B3LYP method with two different basis sets; one can conclude that there is a considerable difference in the interaction energy values. In the case of the D95V basis set one have for the ε interaction energy: −18.10 kcal/mol when the BSSE is taken into account and −21.88 kcal/mol without BSSE correction, while for the D95V++** basis sets the corresponding values are: −13.75 and −14.96, respectively. The BSSE correction is also important, mostly in case of smaller basis set.

Based on the conclusions obtained in the previous, urea dimer conformations investigation, one can point out that the most suitable DFT method is the B3LYP XC functional; namely, if we compare it with the similar MP2 results, nearly the same geometries could be obtained. Moreover, if we take into consideration that the MP2 method slightly overestimates the normal mode vibrational frequencies (especial in the region of 2,500–4,000 cm^{-1}) [74], the B3LYP frequencies show more realistic values as compared to the experimental results. At the same time, some deficiencies can be observed in the case of the far infrared spectra (50–300 cm^{-1}) where the intermolecular normal modes are present. Here, the BHLYP functional showed a better behavior, but we can obtain reasonably good values also with B3LYP functional. Taking all these into considerations and the fact that MP2 calculation has important computational limit, we conclude that in the case of the adenine–thymine DNA base pair (for its geometry see Fig. 7, with the atomic indexes taken from [75]) the most suitable DFT method to calculate harmonic frequency and their anharmonic frequency correction should be the B3LYP functional. The adenine–thymine binary system has 30 atoms and presents a number of 84 normal mode vibrations from which 6 have intermolecular character. Because we are interested in the stretching vibrations (ν_4 and ν_{11}) of those intramolecular covalent bonds where H

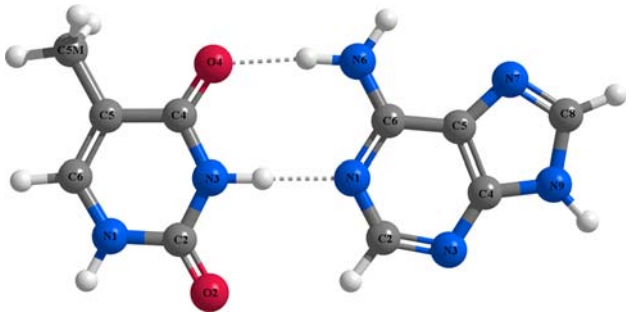

Fig. 7 The equilibrium geometry structure of the adenine–thymine DNA base pair obtained at B3LYP level of theory using the D95V**++ basis set

atoms are involved as well as in those intermolecular vibrations of which motion is situated mostly in the molecular plane (ν_{78}, ν_{80} and ν_{82}), we present only these specific normal mode vibrations. At the same time, one can identify another group of seven normal modes (ν_{41}, ν_{42}, ν_{54}, ν_{60}, ν_{61}, ν_{64}, and ν_{69}) specific for purine and pyrimidine ring vibrational deformation which can significantly disturb the HB vibrations. All these normal mode frequency values in harmonic, BSSE-corrected harmonic, and anharmonic approximations are presented in Table 5, while their normal mode vibrations are shown in ESM Figs. 4 and 5. Considering the BSSE correction, the most affected normal modes are the ν_4 and ν_{11} N–H-stretching vibrations. In these two cases, we found 76.6 cm^{-1} and 429.5 cm^{-1} frequency increasing, respectively. Similar findings can be obtained in case of ν_{78} and ν_{80} intermolecular normal modes, where despite the fact that BSSE effects not show large frequency shifts, they could represent significant corrections when

Table 5 The harmonic (ν) and anharmonic (a) frequencies (cm^{-1}) of some selected intramolecular and intermolecular normal modes in the adenine–thymine DNA base pair, obtained at the B3LYP level of theory and using the D95V basis set (The ν^{int} is the predicted frequency value, considering together the BSSE and anharmonic corrections)

Nr.	ν_{NCP}^{dim}	ν_{CP}^{dim}	a_{NCP}^{dim}	ν^{int}
ν_4	3,329.3	3,405.9	3,048.2	3,124.8
ν_{11}	2,546.7	2,976.2	2,386.3	2,815.8
ν_{41}	1,035.2	1,035.7	1,018.5	1,019.0
ν_{42}	1,031.4	1,024.2	1,014.2	1,007.0
ν_{54}	742.4	745.9	726.9	730.4
ν_{60}	637.7	630.3	632.8	625.4
ν_{61}	607.4	611.1	601.2	604.9
ν_{64}	550.2	550.4	542.0	542.2
ν_{69}	397.2	404.1	388.8	395.7
ν_{78}	119.9	108.2	115.4	103.7
ν_{80}	114.6	101.6	110.3	97.3
ν_{82}	66.8	60.3	58.3	51.8

compared with the uncorrected frequency values. With respect to anharmonic corrections, one can observe that large frequency shifts are obtained in the case of N–H covalent bond stretching vibrations (-281.1 cm^{-1} for ν_4 and -160.4 cm^{-1} for ν_{11}). Considering the ring deformation normal mode vibrations (ν_{41}, ν_{42}, ν_{54}, ν_{60}, ν_{61}, ν_{64}, and ν_{69}), the anharmonic correction is more important than BSSE effects, but even so, their frequency shifts are <20 cm^{-1}. As it was concluded in the previous case of urea dimers, the anharmonic and BSSE corrections could be considered in a very good approximation as additive effects. According to this fact, we present in the fifth column of Table 5, the integral correction (ν^{int}) of the anharmonic and BSSE effects. The results show that in some frequency cases, we have an opposite contribution of frequency shifts, while in some other cases the anharmonic and BSSE collective corrections increase the magnitude of the frequency shift. Analyzing the anharmonic coupling matrix (see ESM Table 6), one observes strong vibrational coupling between ν_{11} intra- and ν_{78} intermolecular normal modes ($x_{11,\ 78} = 8.2$ cm^{-1}) as well as between ν_{11} intra- and ν_{80} intermolecular normal modes ($x_{11,\ 80} = 10.4$ cm^{-1}). The case of this strong coupling could be explained by the same fact that in the case of cyclic urea dimer namely, by the presence of the same molecular plane for both normal modes and the same vibrational direction for ν_{11} and ν_{78} modes. In other cases, there is no significant coupling between the intra- and intermolecular normal modes.

Krishnan et al. [76] showed that comparison of experimentally observed IR spectra of the adenine–thymine base pair with the calculated frequencies was not a simple task. First of all, because the Watson–Crick configuration of A–T base pair is not the most stable isomer conformation [77]. Making a detailed theoretical anharmonic frequency analysis they were able to assign the IR–UV double resonance spectra [78] also to a particular isomer which is not the Watson–Crick structure. A direct experimental assignment of N–H-stretching vibrations in A–T oligomers in condense phase is very difficult because the N–H-stretching vibration spectral region overlaps with the water's O–H-stretching spectral region. Reducing the water content of the A–T oligomers does not solve the problem because they do not adopt a well define structure at extremely low water concentration. To identify and to characterize the N–H-stretching vibrations of A–T base pair oligomers in DNA Heyne et al. [79] use the two-color IR pump-probe technique to overcome the above-mentioned experimental problem. They found that adenine $\nu(NH_2)$ absorbs at 3,215 cm^{-1} and have pronounced anharmonic couplings to the ν(C=O) mode of the thymine and $\delta(NH_2)$ mode of the adenine.

The vibrational relaxation in DNA is a very complicated process. In this paper, only vibrational couplings between the N–H and C=O stretching and different intermolecular (rigid monomer) normal modes, via hydrogen bond, were presented. It is obvious that these normal modes are also influenced by other interactions that are important for the stability and functionality of the DNA. The attractive dispersion effects of DNA base stacking interaction stabilize the base pair's plans. In addition, there is the effect of the phosphate backbone, the ribose or deoxyribose sugars and finally the environment, including counterion (cations) that neutralize the large negative charge due to the phosphate backbone charge. These molecular fragments as well as the interactions between them substantially influence the normal mode vibrations of the base pairs, which also should be considered.

4 Conclusions

When considering the results obtained for two geometry conformations (cyclic and asymmetric) of the urea dimer and the planar geometry configuration of the adenine–thymine DNA base pair, the following conclusions can be drawn:

(1) Considering the HB distance and the intermolecular interaction energy for both cyclic and asymmetric urea dimers, the BLYP, B3LYP, and BHLYP DFT XC functionals gave comparable good results with the second-order Møller–Plesset perturbation theory method. Other DFT functionals such as PBE, HTCH407, and KMLYP underestimate or overestimate the H-bond distances and interaction energies; (2) the empirically corrected DFT functionals which include dispersion effects, the Truhlar's hybrid meta-GGA M06-2x DFT functional or the DF-LMP2 density-fitting local perturbation method are the best theoretical frameworks for describing HBs and intermolecular interaction energies in urea dimers; (3) the best agreement with the MP2 results for normal mode frequency values are obtained for B3LYP and BHLYP functionals; (4) for the N–H and C=O covalent bond stretching vibrations, the anharmonic frequency correction is significant; (5) for the N–H and C=O covalent bond stretching vibrations the BSSE corrections are also significant, but only in the case when the normal mode vibrations are in the intermolecular region; (6) for vibrations which are located in the intermolecular region, the influence of the adjoining molecule (dimer effect) on the vibrational frequency is comparable with the magnitude of the BSSE and anharmonic corrections; (7) analyzing the magnitude of the anharmonic and BSSE corrections obtained for cyclic and asymmetric urea dimers, it was found that their contributions to the harmonic frequency shift could be considered in a very good approximation as an additive effect. (8) The anharmonic coupling between intra- and

intermolecular normal modes is significant only when the motion of normal mode vibrations occur in the same molecular plane (the plane defined by those atoms which move during the vibration) and along the same vibrational direction. (9) Using the classical DFT functionals and applying double-zeta quality basis sets one could obtain a good qualitative description of the anharmonic corrections, but further investigations considering the recently developed DFT or perturbation methods and triple-zeta quality basis sets are needed to obtain a good quantitative description for them.

Acknowledgments This work was supported by the Grant of Romanian Ministry of Education and Research for young researcher, No: CEEX ET-30/08.10.2005, for which A.B. is gratefully acknowledged. A. B. is also thankful for the Grant of German Cancer Research Center (DKFZ) and for its support during his visit to carrying out this work. The author thanks Prof. Dr. Sándor Suhai for helpful comments on the manuscript.

References

1. Rose G, Fleming P, Banavar J, Martin A (2006) PNAS 103:16623
2. Deechongkit S, Nguyen H, Dawson PE, Gruebele M, Kelly JW (2004) Nature 403:101
3. Bende A, Suhai S (2005) Int J Quantum Chem 103(6):841
4. Masunov A, Dannenberg JJ (1999) J Phys Chem A 103:178
5. Masunov A, Dannenberg JJ (2000) J Phys Chem B 104:806
6. Belosludov RV, Li Z-Q, Kawazoe Y (1999) Mol Eng 8:105
7. Skurski P, Simons J (2001) J Chem Phys 115:10731
8. Gavezzotti A (2003) CrystEngComm 5(76):429
9. Caballero-Herrera A, Nilsson L (2006) J Mol Struct (THEOCHEM) 785(2-3):139
10. Shukla A, Isaacs ED, Hamann DR, Platzman PM (2001) Phys Rev B 64:052101
11. Rousseau B, Van Alsenoy C, Keuleers R, Desseyn HO (1998) J Phys Chem A 102:6540
12. Keuleers R, Desseyn HO, Rousseau B, Van Alsenoy C (1999) J Phys Chem A 103:4621
13. Wu K, Snijders JG, Lin C (2002) J Phys Chem B 106:8954
14. Skurski P, Simons J (2001) J Chem Phys 115:8373
15. Skurski P, Simons J (2002) J Chem Phys 116(14):6118
16. Caballero-Herrera A, Nordstrand K, Berndt KD, Nilsson L (2005) Biophys J 89:842
17. Šponer J, Jurečka P, Hobza P (2004) J Am Chem Soc 126:10142
18. Podolyan Y, Nowak MJ, Lapinski L, Leszczynski J (2005) J Mol Struct 744–747:19
19. Guerra CF, Bickelhaupt FM, Snijders JG, Baerends EJ (2000) J Am Chem Soc 122:4117
20. Giese TJ, Sherer EC, Cramer CJ, York DM (2005) J Chem Theory Comput 1:1275
21. Brauer B, Gerber RB, Kabeláč M, Hobza P, Bakker JM, Riziq AGA, de Vries MS (2005) J Phys Chem A 109:6974
22. Bakker JM, Compagnon I, Meijer G, von Helden G, Kabeláč M, Hobza P, de Vries MS (2004) Phys Chem Chem Phys 6:2810
23. Gorb L, Podolyan Y, Dziekonski P, Sokalski WA, Leszczynski J (2004) J Am Chem Soc 126:10119
24. Müller A, Talbot F, Leutwyler S (2002) J Am Chem Soc 124:14486
25. Hobza P, Halvas Z (2000) Chem Rew 100:4253
26. Hobza P, Špirko V (2003) Phys Chem Chem Phys 5:1290
27. Bende A, Vibók Á, Halász GJ, Suhai S (2004) Int J Quantum Chem 99(5):585
28. Wang NX, Venkatesh K, Wilson AK (2006) J Phys Chem A 110:779
29. Simon S, Bertran J, Sodupe M (2001) J Phys Chem A 105:4359
30. Rekik N, Oujia B, Wójcik MJ (2008) Chem Phys 352:65
31. Pelel L, Gerber RB (2008) J Chem Phys 128:165105
32. Watanabe Y, Maeda S, Ohno K (2008) J Chem Phys 129:074315
33. Lundell J, Latajka Z (2008) J Mol Struct 887:179
34. Cremer D (2001) Mol Phys 99:1899
35. Toulouse J, Colonna F, Savin A (2005) J Chem Phys 122:014110
36. Chao SD, Li AH-T (2007) J Phys Chem A 111:9586
37. van Mourik T, Gdanitz RJ (2002) J Chem Phys 116:9620
38. Ireta J, Neugebauer J, Scheffler M (2004) J Phys Chem A 108:5692
39. Johnson ER, Wolkow RA, Di Labio GA (2004) Chem Phys Lett 394:334
40. Kang JK, Musgrave CB (2001) J Chem Phys 115:11040
41. Hamprecht FA, Cohen AJ, Tozer DJ, Handy NC (1998) J Chem Phys 109:6264
42. Boese AD, Doltsinis NL, Handy NC, Sprik M (2000) J Chem Phys 112:1670
43. Boese AD, Handy NC (2001) J Chem Phys 114:5497; see also the supplementary material: EPAPS Document No. E
44. Perdew JP, Burke K, Ernzerhof M (1996) Phys Rev Lett 77:3865
45. Perdew JP, Burke K, Ernzerhof M (1997) Phys Rev Lett 78:1396
46. Jansen HB, Ros P (1969) Chem Phys Lett 3:140; Boys SB, Bernardi F (1970) Mol Phys 19:553
47. Gaussian 03, Revision C.02, 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, Gaussian, Inc., Wallingford CT, 2004
48. Clabo DA Jr, Allen WD, Remington RB, Yamaguchi Y, Schaefer HF III (1988) Chem Phys 123:187
49. Barone V (2005) J Chem Phys 122:014108
50. Dunning TH Jr, Hay PJ (1976) In: Schaefer HF III (ed) Modern theoretical chemistry, vol 3. Plenum, New York, pp 1–28
51. Carbonniere P, Lucca T, Pouchan C, Rega N, Barone V (2005) J Comput Chem 26:384
52. Begue D, Carbonniere P, Pouchan C (2005) J Phys Chem A 109:4611
53. Barone V (2004) J Phys Chem A 108:4146
54. Barone V (2004) Chem Phys Lett 383:528
55. Pulay P (1983) Chem Phys Lett 100:151
56. Hetzer G, Pulay P, Werner H-J (1998) Chem Phys Lett 290:143
57. Schütz M, Hetzer G, Werner H-J (1999) J Chem Phys 111:5691
58. Hetzer G, Schütz M, Stoll H, Werner H-J (2000) J Chem Phys 113:9443
59. Werner H-J, Manby FR, Knowles PJ (2003) J Chem Phys 118:8149
60. MOLPRO, version 2008.1, a package of ab initio programs, Werner H-J, Knowles PJ, Lindh R, Manby FR, Schütz M, Celani

P, Korona T, Mitrushenkov A, Rauhut G, Adler TB, Amos RD, Bernhardsson A, Berning A, Cooper DL, Deegan MJO, Dobbyn AJ, Eckert F, Goll E, Hampel C, Hetzer G, Hrenar T, Knizia G, Köppl C, Liu Y, Lloyd AW, Mata RA, May AJ, McNicholas SJ, Meyer W, Mura ME, Nicklass A, Palmieri P, Pflüger K, Pitzer R, Reiher M, Schumann U, Stoll H, Stone AJ, Tarroni R, Thorsteinsson T, Wang M and Wolf A. http://www.molpro.net
61. Holroyd LF, van Mourik T (2007) Chem Phys Lett 442:42
62. Shields AE, van Mourik T (2007) J Phys Chem A 111:13272
63. Hill JG, Platts JA (2008) Phys Chem Chem Phys 10:2785
64. Schütz M, Rauhut G, Werner H-J (1998) J Phys Chem 102:5197
65. Grimme S (2006) J Comp Chem 27:1787
66. Zhao Y, Truhlar DG (2007) Theor Chem Acc 120:215
67. "NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.1.1", Bylaska EJ, de Jong WA, Govind N, Kowalski K, Straatsma TP, Valiev M, Wang D, Apra E, Windus TL, Hammond J, Nichols P, Hirata S, Hackler MT, Zhao Y, Fan P-D, Harrison RJ, Dupuis M, Smith DMA, Nieplocha J, Tipparaju V, Krishnan M, Vazquez-Mayagoitia A, Wu Q, Van Voorhis T, Auer AA, Nooijen M, Crosby LD, Brown E, Cisneros G, Fann GI, Fruchtl H, Garza J, Hirao K, Kendall R, Nichols JA, Tsemekhman K, Wolinski K, Anchell J, Bernholdt D, Borowski P, Clark T, Clerc D, Dachsel H, Deegan M, Dyall K, Elwood D, Glendening E, Gutowski M, Hess A, Jaffe J, Johnson B, Ju J, Kobayashi R, Kutteh R, Lin Z, Littlefield R, Long X, Meng B, Nakajima T, Niu S, Pollack L, Rosing M, Sandrone G, Stave M, Taylor H, Thomas G, van Lenthe J, Wong A, and Zhang Z, Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA (2009)
68. Korth M, Grimme S (2009) J Chem Theory Comput 5:993
69. Molden 4.7, Schaftenaar G, Noordik JH (2000) J Comput-Aided Mol Design 14:123
70. Gabedit 2.1.11, Allouche A-R, "Gabedit is a free Graphical User Interface for computational chemistry packages." Available from http://gabedit.sourceforge.net/
71. Dunning TH Jr (1989) J Chem Phys 90:1007
72. Kendall RA, Dunning TH Jr, Harrison RJ (1992) J Chem Phys 96:6796
73. Jalkanen KJ, Stephens PJ (1991) J Phys Chem 95:5446
74. Sinha P, Boesch SE, Gu C, Wheeler RA, Wilson AK (2004) J Phys Chem A 108:9213
75. Shui X, McFail-Isom L, Hu GG, Williams LD (1998) Biochem 37:8341
76. Krishnan GM, Kühn O (2007) Chem Phys Lett 435:132
77. Kabeláč M, Hobza P (2001) J Phys Chem B 105:5804
78. Plützer C, Hünig I, Kleinermanns K, Nir E, de Vries MS (2003) Chem Phys Chem 4:838
79. Heyne K, Krishnan GM, Kühn O (2008) J Phys Chem B 112:7909