Non-covalent interactions in extended systems described by the Effective Fragment Potential method: Theory and application to nucleobase oligomers. Supplementary Materials

Debashree Ghosh[1], Dmytro Kosenkov[2], Vitalii Vanovschi[1],

Christopher F. Williams[3], John M. Herbert[3],

Mark S. Gordon[4], Michael W. Schmidt[4],

Lyudmila V. Slipchenko[2] and Anna I. Krylov[1]

[1] Department of Chemistry, University of Southern California, Los Angeles, CA 90089-0482

[2] Department of Chemistry, Purdue University, West Lafayette, IN 47907

[3] Department of Chemistry, Ohio State University, Columbus, OH 43210

[4] Department of Chemistry and Ames Laboratory, Iowa State University, Ames, IA 50011

September 30, 2010

1   Fragment-fragment Coulomb interaction

The equations for the Coulomb fragment-fragment interaction energy are presented below. k and l are the multipole expansion points. q is the charge, μ is the dipole, Θ is the quadrupole, Ω is the octopole, R is the distance between the expansion points k and l, a, b and c are the components of the distance.

Multipole-multipole Coulomb interactions:

$$E_{kl}^{ch-ch} = \frac{q^k q^l}{R_{kl}}$$

$$E_{kl}^{ch-dip} = \frac{q^k \sum_a^{x,y,z} \mu_a^l a}{R_{kl}^3}$$

$$E_{kl}^{ch-quad} = \frac{q^k \sum_a^{x,y,z} \sum_b^{x,y,z} \Theta_{ab}^l ab}{R_{kl}^5}$$

$$E_{kl}^{ch-oct} = \frac{q^k \sum_a^{x,y,z} \sum_b^{x,y,z} \sum_c^{x,y,z} \Omega_{abc}^l abc}{R_{kl}^7}$$

$$E_{kl}^{dip-dip} = \frac{\sum_a^{x,y,z} \mu_a^k \mu_a^l}{R_{kl}^3} - 3\frac{\sum_a^{x,y,z} \sum_b^{x,y,z} \mu_a^k \mu_b^l ab}{R_{kl}^5}$$

$$E_{kl}^{dip-quad} = -2\frac{\sum_a^{x,y,z} \sum_b^{x,y,z} \Theta_{ab}^k \mu_a^l b}{R_{kl}^5} + 5\frac{\sum_a^{x,y,z} \sum_b^{x,y,z} \Theta_{ab}^k ab \sum_c^{x,y,z} \mu_c^l c}{R_{kl}^7}$$

$$E_{kl}^{quad-quad} = 2\frac{\sum_a^{x,y,z} \sum_b^{x,y,z} \Theta_{ab}^k \Theta_{ab}^l}{3R_{kl}^5}$$

$$-20\frac{\sum_a^{x,y,z} \sum_b^{x,y,z} \Theta_{ab}^k b \sum_c^{x,y,z} \Theta_{ac}^l c}{3R_{kl}^7} + 35\frac{\sum_a^{x,y,z} \sum_b^{x,y,z} \Theta_{ab}^k ab}{3R_{kl}^9}$$

(1)

Multipole-nuclei terms:

$$E_{Il}^{nuc-ch} = \frac{Z^I q^l}{r_{kl}}$$

$$E_{Il}^{nuc-dip} = \frac{Z^I \sum_a^{x,y,z} \mu_a^l a}{R_{kl}^3}$$

$$E_{Il}^{nuc-quad} = \frac{Z^I \sum_a^{x,y,z} \sum_b^{x,y,z} \Theta_{ab}^l ab}{R_{kl}^5}$$

$$E_{Il}^{nuc-oct} = \frac{Z^I \sum_a^{x,y,z} \sum_b^{x,y,z} \sum_c^{x,y,z} \Omega_{abc}^l abc}{R_{kl}^7}$$

(2)

2 Buckingham multipoles

2.1 The conversion from the spherical to the Buckingham multipoles

The Buckingham charge q in terms of the spherical multipole Q with $l = 0$ angular momentum:

$$q = Q_{0,0} \tag{3}$$

The Buckingham dipole μ in terms of the spherical multipole Q with $l = 1$ angular momen-

tum:

$$\mu_x = -2Q_{11}$$

$$\mu_y = 2Q_{1\bar{1}} \tag{4}$$

$$\mu_z = Q_{10}$$

The Buckingham quadrupole Θ in terms of the spherical multipole Q with $l = 2$ angular momentum:

$$\Theta_{xx} = 6Q_{22} - Q_{20}$$

$$\Theta_{xy} = -6Q_{2\bar{2}}$$

$$\Theta_{yy} = -6Q_{22} - Q_{20}$$

$$\Theta_{xz} = -3Q_{21} \tag{5}$$

$$\Theta_{yz} = 3Q_{2\bar{1}}$$

$$\Theta_{zz} = 2Q_{20}$$

The Buckingham octopole Ω in terms of the spherical multipole Q with $l = 3$ angular momentum:

$$\Omega_{xxx} = -30Q_{33} + 6Q_{31}$$

$$\Omega_{xxy} = 30Q_{3\bar{3}} - 2Q_{3\bar{1}}$$

$$\Omega_{xxz} = -3Q_{30} + 10Q_{32}$$

$$\Omega_{xyy} = 30Q_{33} + 6Q_{31}$$

$$\Omega_{xyz} = -10Q_{3\bar{2}}$$

$$\Omega_{xzz} = -8Q_{31} \tag{6}$$

$$\Omega_{yyy} = -30Q_{3\bar{3}} - 6Q_{3\bar{1}}$$

$$\Omega_{yyz} = -3Q_{30} - 10Q_{32}$$

$$\Omega_{yzz} = 8Q_{3\bar{1}}$$

$$\Omega_{zzz} = 6Q_{30}$$

2.2 The conversion from the Cartesian to the Buckingham multipoles

.

The Buckingham charge q in terms of the Cartesian charge q':

$$q = q'$$

(7)

The Buckingham dipole μ in terms of the Cartesian dipole μ':

$$\mu_x = \mu'_x$$
$$\mu_y = \mu'_y$$
$$\mu_z = \mu'_z$$

(8)

The Buckingham quadrupole Θ in terms of the Cartesian quadrupole Θ':

$$halftrace = \frac{\Theta'_{xx} + \Theta'_{yy} + \Theta'_{zz}}{2}$$
$$\Theta_{xx} = \frac{3}{2}\Theta'_{xx} - halftrace$$
$$\Theta_{xy} = \frac{3}{2}\Theta'_{xy}$$
$$\Theta_{yy} = \frac{3}{2}\Theta'_{yy} - halftrace$$
$$\Theta_{xz} = \frac{3}{2}\Theta'_{xz}$$
$$\Theta_{yz} = \frac{3}{2}\Theta'_{yz}$$
$$\Theta_{zz} = \frac{3}{2}\Theta'_{zz} - halftrace$$

(9)

The Buckingham octopole Ω in terms of the Cartesian octopole Ω':

$$traceX = \Omega'_{xxx} + \Omega'_{xyy} + \Omega'_{xzz}$$

$$traceY = \Omega'_{xxy} + \Omega'_{yyy} + \Omega'_{yzz}$$

$$traceZ = \Omega'_{xxz} + \Omega'_{yyz} + \Omega'_{zzz}$$

$$\Omega_{xxx} = \frac{5}{2}\Omega'_{xxx} - \frac{3}{2}traceX$$

$$\Omega_{xxy} = \frac{5}{2}\Omega'_{xxy} - \frac{3}{2}traceY$$

$$\Omega_{xxz} = \frac{5}{2}\Omega'_{xxz} - \frac{3}{2}traceZ$$

$$\Omega_{xyy} = \frac{5}{2}\Omega'_{xyy} - \frac{1}{2}traceY$$

$$\Omega_{xyz} = \frac{5}{2}\Omega'_{xyz} - \frac{1}{2}traceZ$$

$$\Omega_{xzz} = \frac{5}{2}\Omega'_{xzz} - \frac{1}{2}traceX$$

$$\Omega_{yyy} = \frac{5}{2}\Omega'_{yyy} - \frac{1}{2}traceX$$

$$\Omega_{yyz} = \frac{5}{2}\Omega'_{yyz} - \frac{1}{2}traceZ$$

$$\Omega_{yzz} = \frac{5}{2}\Omega'_{yzz} - \frac{1}{2}traceY$$

$$\Omega_{zzz} = \frac{5}{2}\Omega'_{zzz}$$

(10)

3   Effective fragment library of common solvents

Currently, the library includes 12 common organic solvents shown in Figs. 1 and 2. Unique labels were assigned to each molecule in the library (given in parentheses): acetone (ACETONE_L), carbon tetrachloride (CCL4_L), dichloromethane (DCM_L), methane (METHANE_L), methanol (METHANOL_L), ammonia (AMMONIA_L), acetonitrile (ACETONITRILE_L), water (WATER_L), dimethyl sulfoxide (DMSO_L), benzene (BENZENE_L), phenol (PHENOL_L), and toluene (TOLUENE_L). The geometries of the molecules were optimized by MP2/cc-pVTZ. To generate the electrostatic multipoles and electrostatic screening parameters, analytic DMA procedure was used, with 6-31+G* basis for non-aromatic compounds and 6-31G* for aromatic compounds. The rest of the potential, i.e., static and dynamic polarizability tensors, wave function, Fock matrix, etc., were obtained with 6-311++G(3df,2p) basis set.

The library also includes DNA bases and different tautomers of guanine and cytosine, shown

in Figs. 3-5. The geometries of the bases were optimized with RI-MP2/cc-pVTZ, and EFP parameters were generated with 6-31G* basis for electrostatic multipoles (analytic DMA procedure) and electrostatic screening parameters and 6-311++G(3df,2p) basis for all other components of the potential.

The EFP parameters for other molecules can be generated using the *GAMESS* program keyword RUNTYP=MAKEFP (see the GAMESS manual for details). The GAMESS EFP parameters can be converted to *Q-CHEM* library format using a script (see the *Q-CHEM* manual for details).

We also developed a set of scripts allowing conversion of the geometry of the system consisting of molecular fragments appearing in the library from the PDB file format to the EFP input format (see the Q-CHEM manual). For the correct alignment and recognition of the geometries of molecules, the labels (tags) of atoms in the input file should correspond to those in the library. The atom numbering and atomic labels (tags) of the fragments are shown in Figs. 1-2.
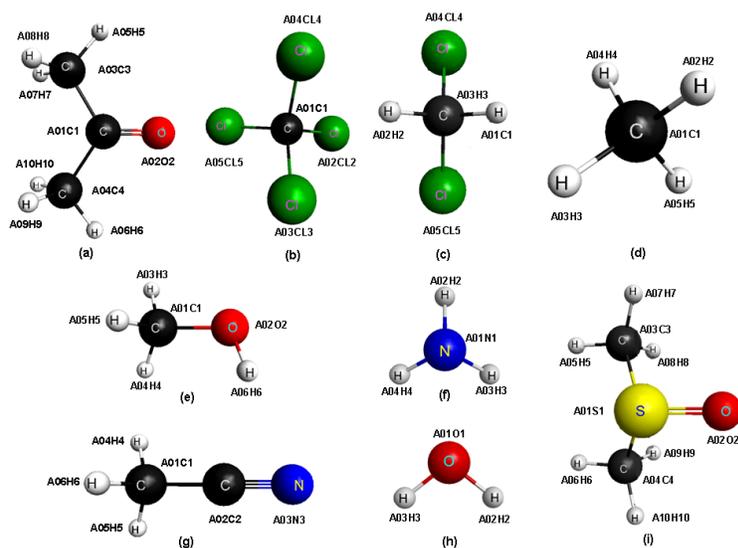

Figure 1: Standard EFP fragments: (a) acetone (ACETONE_L), (b) carbon tetrachloride (CCL4_L), (c) dichloromethane (DCM_L), (d) methane (METHANE_L), (e) methanol (METHANOL_L), (f) ammonia (AMMONIA_L), (g), acetonitrile (ACETONITRILE_L), (h) water (WATER_L), (i) dimethyl sulfoxide (DMSO_L).
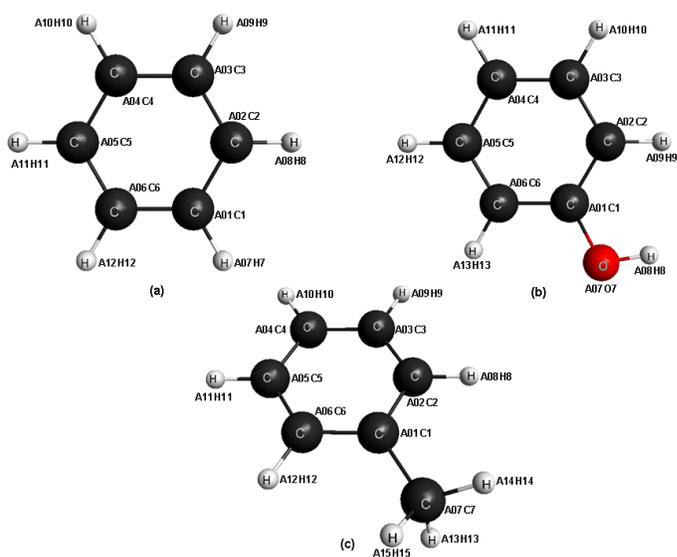
Figure 2: Standard EFP fragments: (a) benzene (BENZENE_L), (b) phenol (PHENOL_L), and (c) toluene (TOLUENE_L).


(a)   (b)
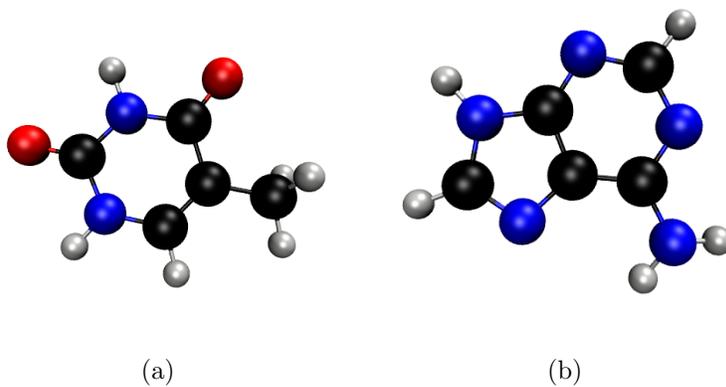
Figure 3: Standard EFP fragments: (a) thymine (THYMINE_L) and (b) adenine (ADE NINE_L).

(a)

(b)

(c)

(d)

(e)
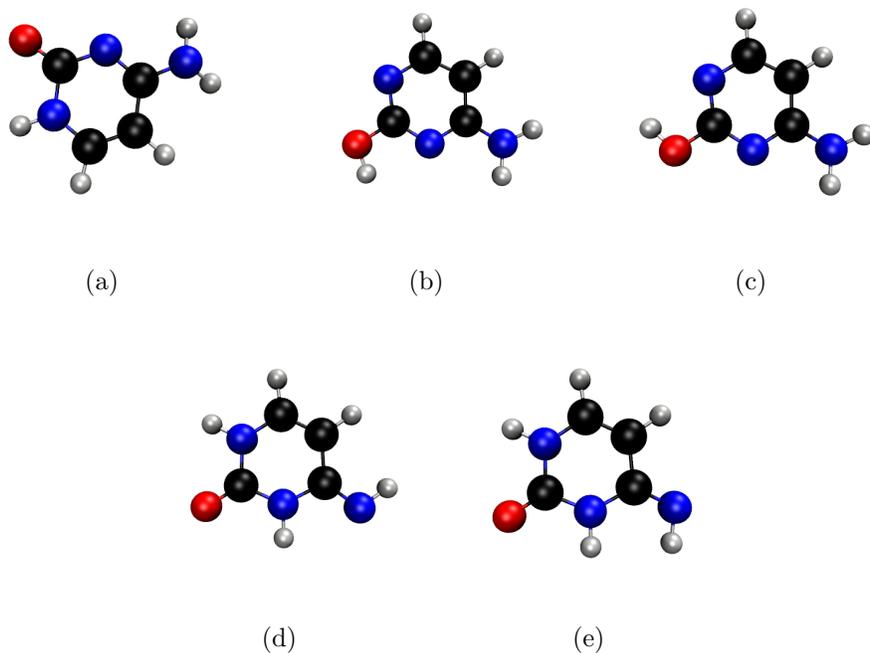
Figure 4: Standard EFP fragments: cytosine tautomers (a) C1 (CYTOSINE_C1_L), (b) C2a (CYTOSINE_C2A_L), (c) C2b (CYTOSINE_C2B_L), (d) C3a (CYTOSINE_C3A_L), and (e) C3b (CYTOSINE_C3B_L).


(a)

(b)

(c)

(d)

(e)
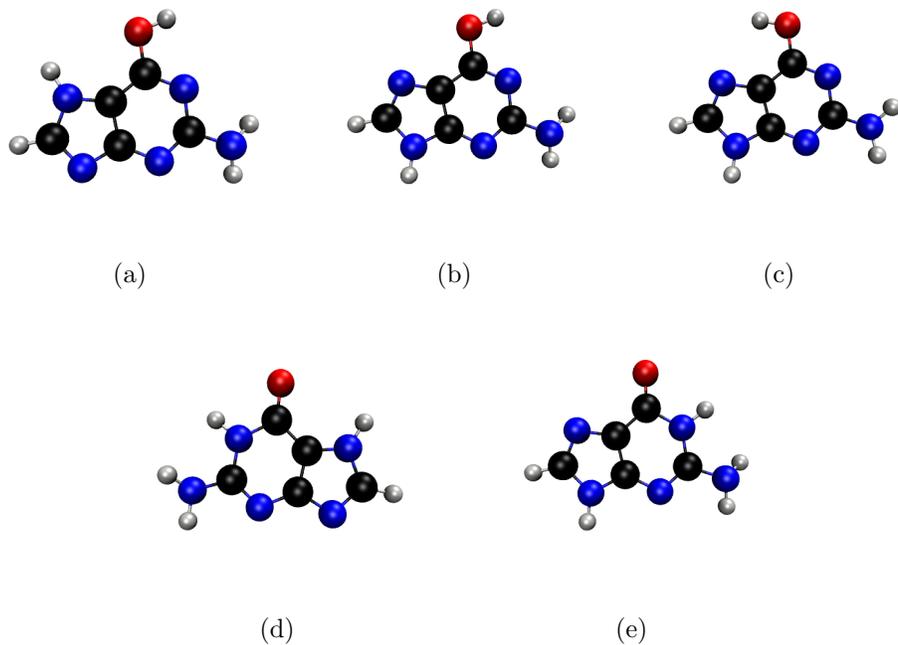
Figure 5: Standard EFP fragments: guanine tautomers (a) enol N7 (GUANINE_EN7_L), (b) enol N9 (GUANINE_EN9_L), (c) enol N9RN7 (GUANINE_EN9RN7_L), (d) keton N7 (GUANINE_KN7_L), and (e) keton N9 (GUANINE_KN9_L).

4   Adenine and thymine dimers: Calibration against ab initio calculations

We use the dimers to assess the accuracy of the EFP-EFP interaction energies. More detailed analysis of the interactions of heterocyclic aromatic compounds including DNA bases will appear in a separate publication.[1]

The $\pi - \pi$ non-bonding interactions are difficult to compute because they are dominated by the dispersion energy which requires correlated *ab initio* methods for accurate description.

Previous theoretical studies[2–4] have shown that there are numerous isomers of the adenine-adenine (AA), thymine-thymine (TT) and adenine-thymine (AT) dimers. The structures of adenine and thymine dimers used to compare the performance of EFP against *ab initio* methods were taken from Ref. 4 [optimized with ωB97X-D/6-31+G(d,p) [5]]. The lowest-energy AA, TT and AT structures were used.

Table 1: Binding energy (D_e, not ZPE corrected) of the AA stacked and H-bonded dimers at various levels of theory.

Method	Stacked, kcal/mol	H-bonded, kcal/mol
RI-MP2/aug-cc-pVTZ [a]	11.10	14.20
MP2/6-31G*[b]	8.83	11.55
MP2/6-31+G(d,p)[c]	10.66	20.78
MP2/6-31++G(2df,2pd)[c]	10.55	19.85
EFP	9.03	16.03

[a] From Ref. 6.

[b] From Ref. 7.

[c] From Ref. 4.

Tables 1, 2 and 3 compare the binding energies of the most stable stacked and H-bonded adenine and thymine dimers. For the stacked dimers, the EFP interaction energies are in good agreement with the energies obtained by high level *ab initio* calculations (the errors are ≈ 2 kcal/mol) when compared to the MP2 calculations carried out with the same structures and the 6-31+G(d,p) and 6-311++G(2df,3p) basis sets. The EFP interaction energies are consis-

Figure 6: Optimized structures of the stacked and hydrogen bonded nucleic acid base dimers.
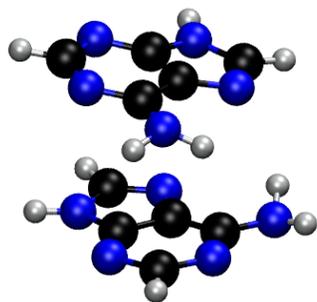
(a) Stacked adenine dimer
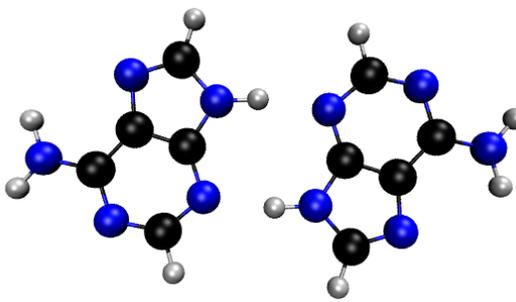
(b) H-bonded adenine dimer
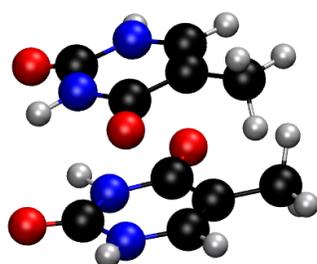
(c) Stacked thymine dimer
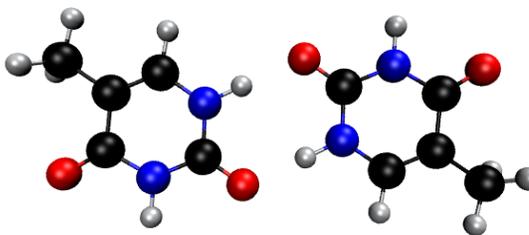
(d) H-bonded thymine dimer
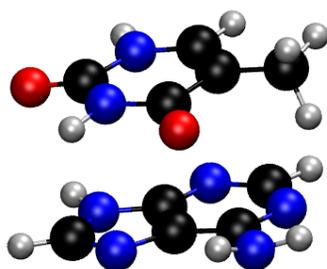
(e) Stacked adenine-thymine dimer
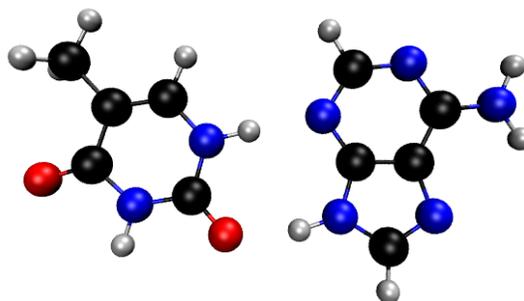
(f) H-bonded adenine-thymine dimer

tently under-estimated, primarily due to the error in polarization component. The polarization energies can be corrected by using more accurate DFT-based polarizabilities.

The EFP energy calculations are few orders of magnitude faster than the MP2 calculations and take only a few seconds. The BSSE corrections due to incomplete basis sets used in the MP2 calculations are expected to be small since the difference between the MP2 energies with

Table 2: Binding energy (D_e, not ZPE corrected) of the TT stacked and H-bonded dimers at various levels of theory.

Method	Stacked, kcal/mol	H-bonded, kcal/mol
MP2 [a]	-	10.6
MP2/6-31+G(d,p) [b]	11.49	19.28
MP2/6-31++G(2df,2pd) [b]	11.59	19.33
EFP	9.92	17.18

[a] From Ref. 8. Note that the structures are not identical to the ones used in the EFP calculations.

[b] From Ref. 4.

Table 3: Binding energy (D_e, not ZPE corrected) of the AT stacked H-bonded dimers at various levels of theory.

Method	Stacked, kcal/mol	H-bonded, kcal/mol
RI-MP2/aug-cc-pVQZ [a]	14.78	16.53
MP2 [b]	12.4	13.3
MP2/6-31+G(d,p) [c]	12.39	19.80
MP2/6-311++G(2df,2pd) [c]	12.38	19.22
EFP	10.71	16.48

[a] From Ref. 6.

[b] From Ref. 8.

[c] From Ref. 4.

different number of diffuse function is only a tenth of a kcal/mol. Thus, the EFP energies agree with the MP2 energies within the error bars of MP2. The tables also compare the interaction energies with those from the previous studies, Refs. 6–8, which also used slightly different structures. The EFP binding energies are comparable to the energies computed by these high-level calculations.

Table 4 shows the breakdown of the EFP fragment-fragment energy into Coulomb, exchange-repulsion, polarization and dispersion energies for the stacked adenines and thymines. We notice

Table 4: Different components of the interaction energies (in percentage) for the optimized stacked AA and TT dimers at their equilibrium geometries.

Energy component	AA dimer	TT dimer
Coulomb energy	58.18	105.07
Exchange-Repulsion energy	-142.70	-196.85
Polarization energy	5.40	6.81
Dispersion energy	179.12	184.98

that the major attractive contribution comes from the dispersion energy and is opposed by the exchange-repulsion energy. The contribution of the Coulomb energy is higher for the thymine dimer than the adenine dimer since the thymine molecule is more polar and the polar CO and NH bonds are aligned in the lowest energy structure maximizing favorable dipolar interactions (interestingly, this is different in B-DNA structure). Polarization contribution is relatively small for both dimers.

5 Effect of h-twist on the components of interaction energy in thymine dimer

Figure 7: The effect of the h-twist on the components of the fragment interaction energy for the thymine dimer.
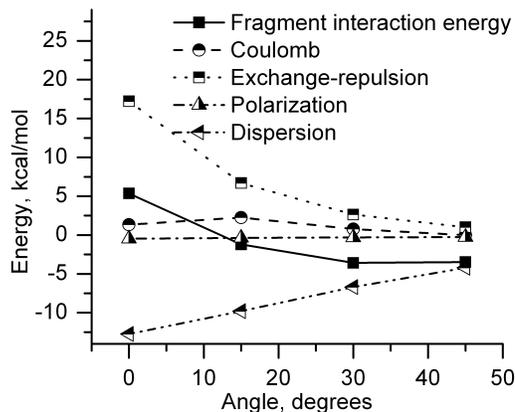
From Fig. 7, we notice that the fragment interaction energy of thymine dimers decrease monotonically when the h-twist angle increases from 0° to 45°. The major effect is from

the decrease in exchange-repulsion due to the thymine molecules moving further from each other. However, the total stabilization energy is very small due to the opposing effect of the dispersion and very small change in Coulomb components. In the case of the adenine dimers this is clearly not the case as we can notice on comparing the 0° and 38° (real B-DNA) h-twist (the change in fragment interaction energy changes from -0.4 kcal/mol to -6.7 kcal/mol where there is a significant decrease in Coulomb energy leading to the total stabilization energy being substantial).

6 Far-field interactions and their dependence on the size of the oligomers

In this section we analyze non-additive component of far-field interactions [see main text for the definition of total pairwise component of the interaction energy, $E_2(\text{total})$, and many-body or non-additive contributions (E_{mb} and Δ)].

The total two-body energy [$E_2(\text{total})$ defined by Eq. (39) in the main body of the paper] of the N oligomers can be partitioned into the nearest-neighbor pair-wise interaction energy [$E_2(1,2)$, $E_2(2,3)$, ...], next-nearest-neighbor pair-wise interactions [$E_2(1,3)$, $E_2(2,4)$, ...], and so on, as follows:

$$
\begin{aligned}
E_2(\text{total}) &= [E_2(1,2) + E_2(2,3) + \cdots + E_2(N-1,N)] \\
&\quad + [E_2(1,3) + \cdots + E_2(N-2,N)] \\
&\quad + \cdots + [E_2(1,N-1) + E_2(2,N)] + E_2(1,N).
\end{aligned}
\tag{11}
$$

Thus, one can simplify the expression for the many-body (or non-additive) energy and the corresponding fraction by approximating the total pairwise energy $E_2(\text{total})$ by the energies of the neighboring dimers only:

$$
\begin{aligned}
E_{\text{non-add}}^{1-2} &= E_{\text{total}} - [E_2(1,2) + E_2(2,3) + \cdots + E_2(N-1,N)] \\
&= E_{\text{total}} - E_2(\text{total}; N-1, N) \\
&= E_{\text{total}} - (n-1) * E_2^{1-2},
\end{aligned}
\tag{12}
$$

$$
\Delta^{1-2} = \frac{E_{\text{non-add}}^{1-2}}{E_{\text{total}}} \times 100\%,
\tag{13}
$$

13

where E_2^{1-2} denotes the nearest-neighbor or direct two-body interactions such as $E_2(1,2)$, $E_2(2,3)$ etc. In this formulation, all nearest-neighbor interactions are assumed to be equal.

Figure 8: Δ^{1-2} and Δ^{1-2-3} approximations to the non-additive part of the interaction energy components for stacked adenine (top), thymine (middle), and adenine-thymine (bottom) oligomers. The Coulomb component in the thymine plot is omitted due to its high values (see Table 5 for more details).
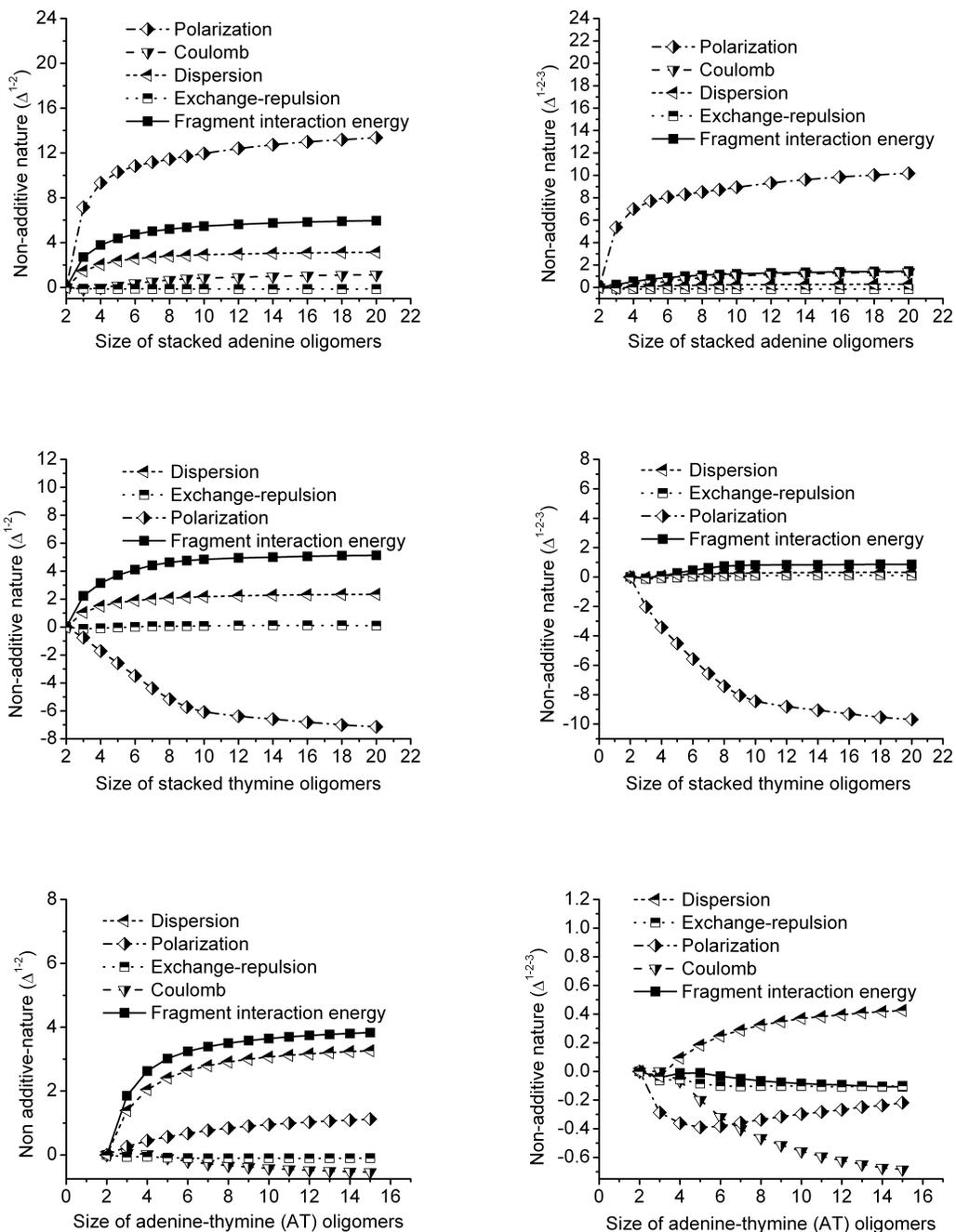
A more accurate approximation to non-additive interactions is is obtained by representing the total pairwise energy of the system as a sum of energies of nearest neighbors and next-nearest neighbors:

$$
\begin{aligned}
E_{\text{non-add}}^{1-2-3} &= E_{\text{total}} - [E_2(1,2) + E_2(2,3) + \cdots + E_2(N-1,N)] - \\
&\quad [E_2(1,3) + E_2(2,4) + \cdots E_2(N-2,N)] \\
&= E_{\text{total}} - E_2(\text{total}; N-1, N) - E_2(\text{total}; N-2, N) \\
&= E_{\text{total}} - (n-1) * E_2^{1-2} - (n-2) * E_2^{1-3},
\end{aligned} \tag{14}
$$

where E_2^{1-2} denotes the nearest-neighbor two-body interactions, such as $E_2(1,2)$, $E_2(2,3)$,..., and E_2^{1-3} denotes the next-nearest neighbor (geminal) two-body interactions such as $E_2(1,3)$, $E_2(2,4)$, ... Here again all nearest-neighbor interactions are treated as equal, and all next-nearest neighbor interactions are equal as well. The corresponding fraction of the non-additive interaction energy is given by:

$$
\Delta^{1-2-3} = \frac{E_{\text{non-add}}^{1-2-3}}{E_{\text{total}}} \times 100\%. \tag{15}
$$

Extrapolating this approach up to $1-n$ interacting dimers, in the limit of $n \to N$, one obtains the total (exact) pairwise energy $E_2(\text{total})$ of the system. In this limit, the fraction of the non-additive energy Δ is a fraction of the intrinsically non-additive contributions to the total interaction energy of the system. However, in the case of approximating the total pairwise energy by the near-field energies, the fraction of the non-additive energy differs from this limiting value by the fraction of the far-field pairwise energies. Fig. 8 and Table 5 analyze the magnitude and dependence of the far-field contributions of the pairwise energies on the size of the oligomer. Both total interaction energies and the energy components are analyzed.

Fig. 8 shows the percentage of the approximate non-additive energies, Δ^{1-2} and Δ^{1-2-3}, versus the oligomer size. Generally, the values of Δ^{1-2} for the total interaction energies increase with the size of the oligomers and approach a constant at around 10-mers ($\approx 4\%$). By analyzing the trends for the individual components in adenine, the far-field part is highest for polarization ($\approx 15\%$) followed by dispersion ($\approx 3\%$) and electrostatics ($\approx 2\%$) energy. The non-additive fraction of exchange-repulsion is nearly zero for all oligomers. The large values of far-field contribution in polarization can be easily rationalized since the polarization is the

only intrinsically non-additive term in the EFP interaction energy. All other components of the interaction energy are exactly pairwise.

By using a more accurate representation of the non-additive interaction energy, Δ^{1-2-3}, we observe that the percentage of the non-additive part in the polarization is similar to that in Δ^{1-2}, while the far-field parts in the Coulomb and dispersion energies are significantly reduced. Consequently, the far-field part of the total interaction energy also decreases to 1% when using Δ^{1-2-3}. Thus, at the limit of considering *all* of the pair-wise interactions, the only significant contribution of the far-field part of the energy is from the intrinsically non-additive polarization. Therefore, the polarization component of the interaction energy cannot be accurately approximated by a sum of all pair-wise interactions and that three and higher body interactions are required to correctly describe the polarization in extended systems.

Another important result is that for all of the energy components, the far-field fraction saturates at 8-10-mers, whereas in the case of polarization energy we observe small changes in the non-additive percentage even in 20-mers. Thus, not only does the polarization have a significant non-additive character, but it also continues to vary with the system size even for large oligomers.

Table 5: Asymptotic limit of non-pairwise components of the interaction energies (kcal/mol) in thymine oligomers at various levels of approximation. The $\Delta^{1-2-3-4}$ and $\Delta^{1-2-3-4-5}$ are defined similarly to Δ^{1-2-3} and take into account the interactions between $n \leftrightarrow (n+3)$ and $n \leftrightarrow (n+4)$ fragments, respectively.

Approximation	Frag. inter-action energy	Coulomb Energy	Exch-repulsion Energy	Polarization Energy	Dispersion Energy
Δ^{1-2}	5.135	-92.373	0.099	-7.141	2.342
Δ^{1-2-3}	0.859	-40.560	0.097	-9.688	0.322
$\Delta^{1-2-3-4}$	0.126	-14.523	0.097	-9.950	0.115
$\Delta^{1-2-3-4-5}$	-0.210	-0.210	0.097	-10.010	0.069

Although the thymine oligomers show trends similar to those of adenine stacks, we had to employ higher-level approximations, $\Delta^{1-2-3-4}$ and $\Delta^{1-2-3-4-5}$, to observe that polarization is indeed the only component with a significant non-pairwise character. The asymptotic values of non-pairwise contributions in thymine oligomers at various levels of approximation are tab-

ulated in Table 5 demonstrating that the fraction of the far-field components of the Coulomb term in thymine is very high (-92% for the nearest neighbors and -40% for the next-nearest neighbors). This may be either due to a more polar nature of thymine or because the magnitude of the Coulomb contribution in the thymine stack is very small (2 kcal/mol), such that small variations of the far-field components result in large ratios.

The adenine-thymine oligomers exhibit trends that are rather different from the pure stacks of individual NABs. Fig. 8 (bottom) shows that the different energy components do not follow a simple pattern. The reason is again in the varying number of different kinds of interactions in the oligomers that cannot be represented by a simple summation of the dimer interactions. However, the far-field part of the total interaction energy is smaller ($\Delta^{1-2-3} \approx 0.1\%$) than in the pure stacks. Therefore, although the individual energy components cannot be accurately approximated as a sum of the (AT) dimer energies, the sum of the total dimer interaction energies agrees well with the total oligomer energy.

7   Geometries of the oligomers used in the Q-Chem format

Geometry of ADENINE monomer

```
C -0.1908012267 -0.5247948651 0.0010972008
N -1.2026293848 -1.4577534979 -0.0079312871
C -2.2987002719 -0.7184880661 -0.0057519465
H -3.3028846013 -1.1084590586 -0.0109747137
N -2.0515535931 0.6267580071 0.0043227613
H -2.7297797944 1.3702392898 0.0069546605
C -0.6864481258 0.7805989093 0.0059297778
N 0.0255173477 1.9117076406 0.0058001875
C 1.3330378539 1.6496441008 -0.0056445079
H 1.9891566180 2.5098480287 -0.0112943846
N 1.9551897996 0.4543447430 -0.0121773802
C 1.2082292941 -0.6519949127 -0.0021319337
```

```
N 1.8167745041 -1.8655354663 0.0447991624
H 1.2592318326 -2.6665519625 -0.1952761804
H 2.7925515489 -1.8830311981 -0.1957082713
```

Geometry of THYMINE monomer

```
N -0.6310683796 1.0083732671 0.0000000000
H -0.9566160144 1.9653029408 0.0000000000
C -1.6219633191 0.0462759542 0.0000000000
O -2.8118960228 0.2945417514 0.0000000000
N -1.1112695344 -1.2365310932 0.0000000000
H -1.7990850227 -1.9709039294 0.0000000000
C 0.2323289640 -1.5249962904 0.0000000000
H 0.4798513273 -2.5770009882 0.0000000000
C 1.1832459322 -0.5662330606 0.0000000000
C 0.7558354625 0.8285886265 0.0000000000
O 1.5089252931 1.7895081368 0.0000000000
C 2.6487124384 -0.8489169744 0.0000000000
H 2.8384407233 -1.9201204900 0.0000000000
H 3.1226017923 -0.4057962348 -0.8743590000
H 3.1226017923 -0.4057962348 0.8743590000
```

Table 6: $(A)_n$ stacked geometry. Distances in Å and angles in degrees.

Fragment	x	y	z	α	β	γ
ADENINE	-1.37701480	1.41334560	3.40415522	2.504756	3.030456	-3.010720
ADENINE	-1.34215671	0.11877860	6.80424961	-2.807868	3.052053	-2.714643
ADENINE	-0.41751883	-0.77722236	10.20846312	-1.771379	3.037971	-2.353078
ADENINE	0.95164612	-0.80560772	13.58086518	-1.002946	3.002090	-2.260541
ADENINE	2.12373124	0.11994004	16.89998448	-0.414371	2.967717	-2.349515
ADENINE	2.54285947	1.66454159	20.16828861	0.099749	2.945541	-2.514201
ADENINE	1.98597690	3.22046085	23.41154784	0.591815	2.941133	-2.702216
ADENINE	0.66221268	4.17265177	26.66487722	1.089180	2.953924	-2.884174
ADENINE	-0.88298871	4.17604688	29.96032988	1.631204	2.982937	-3.021227
ADENINE	-2.00275716	3.30201955	33.31028437	2.281536	3.019787	-3.047110
ADENINE	-2.24318220	2.00911812	36.70470978	-3.099211	3.049142	-2.820071
ADENINE	-1.53354352	0.94299320	40.11155793	-2.028626	3.045187	-2.424217
ADENINE	-0.22682419	0.64898951	43.49452636	-1.196459	3.012508	-2.267159
ADENINE	1.06343837	1.32778425	46.82917983	-0.572405	2.975934	-2.319499
ADENINE	1.73088413	2.75580930	50.11010504	-0.038791	2.950240	-2.465619
ADENINE	1.44129178	4.37491195	53.35659005	0.455115	2.940073	-2.651252
ADENINE	0.28552017	5.54305968	56.60337342	0.946251	2.947950	-2.839768
ADENINE	-1.26509553	5.82079118	59.88521282	1.473359	2.974337	-2.992268
ADENINE	-2.56069052	5.15566424	63.22167227	2.090721	3.011002	-3.051846
ADENINE	-3.06880952	3.91780863	66.60533711	2.904954	3.043882	-2.912643

Table 7: $(T)_n$ stacked geometry. Distances in Å and angles in degrees.

Fragment	x	y	z	α	β	γ
THYMINE	2.70535852	-2.93404777	3.44635270	2.441407	0.178420	-2.022511
THYMINE	4.56142619	-0.71757439	6.71838984	2.998602	0.208194	-1.901845
THYMINE	4.70520979	2.26146563	9.90786313	-2.769206	0.221195	-1.736205
THYMINE	3.03498354	4.75889720	13.07509946	-2.260233	0.215322	-1.565844
THYMINE	0.25045021	5.74904154	16.28875236	-1.724254	0.191477	-1.421839
THYMINE	-2.45674990	4.86589053	19.59884946	-1.119599	0.156092	-1.349276
THYMINE	-3.92891825	2.57328464	23.01098883	-0.348363	0.121228	-1.445012
THYMINE	-3.55302691	-0.04205654	26.48719059	0.641755	0.109196	-1.760568
THYMINE	-1.53428705	-1.75176130	29.96011645	1.564052	0.131585	-2.008705
THYMINE	1.19931770	-1.72877957	33.36419682	2.271715	0.168005	-2.040369
THYMINE	3.40407717	0.09274343	36.66452153	2.846568	0.200428	-1.936760
THYMINE	4.06658674	2.97947326	39.87235027	-2.910786	0.218334	-1.782892
THYMINE	2.85712855	5.73255638	43.03872760	-2.402415	0.219068	-1.610790
THYMINE	0.27441268	7.20892763	46.23348818	-1.879513	0.199625	-1.454431
THYMINE	-2.58188696	6.83103172	49.51440636	-1.299392	0.166294	-1.356310
THYMINE	-4.48742799	4.83816185	52.90102652	-0.580295	0.130046	-1.399119
THYMINE	-4.63711627	2.18501368	56.36573448	0.359610	0.109313	-1.664380
THYMINE	-3.00493798	0.11686342	59.84579800	1.330966	0.122055	-1.962099
THYMINE	-0.34704645	-0.38111436	63.27495340	2.093966	0.158537	-2.049785
THYMINE	2.12340487	0.98566531	66.60507754	2.697723	0.192878	-1.976231

Table 8: $(A - T)_n$ stacked geometry. Distances in Å and angles in degrees.

Fragment	x	y	z	α	β	γ
ADENINE	-1.37701480	1.41334560	3.40415522	2.504756	3.030456	-3.010720
ADENINE	-1.34215671	0.11877860	6.80424961	-2.807868	3.052053	-2.714643
ADENINE	-0.41751883	-0.77722236	10.20846312	-1.771379	3.037971	-2.353078
ADENINE	0.95164612	-0.80560772	13.58086518	-1.002946	3.002090	-2.260541
ADENINE	2.12373124	0.11994004	16.89998448	-0.414371	2.967717	-2.349515
ADENINE	2.54285947	1.66454159	20.16828861	0.099749	2.945541	-2.514201
ADENINE	1.98597690	3.22046085	23.41154784	0.591815	2.941133	-2.702216
ADENINE	0.66221268	4.17265177	26.66487722	1.089180	2.953924	-2.884174
ADENINE	-0.88298871	4.17604688	29.96032988	1.631204	2.982937	-3.021227
ADENINE	-2.00275716	3.30201955	33.31028437	2.281536	3.019787	-3.047110
ADENINE	-2.24318220	2.00911812	36.70470978	-3.099211	3.049142	-2.820071
ADENINE	-1.53354352	0.94299320	40.11155793	-2.028626	3.045187	-2.424217
ADENINE	-0.22682419	0.64898951	43.49452636	-1.196459	3.012508	-2.267159
ADENINE	1.06343837	1.32778425	46.82917983	-0.572405	2.975934	-2.319499
ADENINE	1.73088413	2.75580930	50.11010504	-0.038791	2.950240	-2.465619
THYMINE	2.70535852	-2.93404777	3.44635270	2.441407	0.178420	-2.022511
THYMINE	4.56142619	-0.71757439	6.71838984	2.998602	0.208194	-1.901845
THYMINE	4.70520979	2.26146563	9.90786313	-2.769206	0.221195	-1.736205
THYMINE	3.03498354	4.75889720	13.07509946	-2.260233	0.215322	-1.565844
THYMINE	0.25045021	5.74904154	16.28875236	-1.724254	0.191477	-1.421839
THYMINE	-2.45674990	4.86589053	19.59884946	-1.119599	0.156092	-1.349276
THYMINE	-3.92891825	2.57328464	23.01098883	-0.348363	0.121228	-1.445012
THYMINE	-3.55302691	-0.04205654	26.48719059	0.641755	0.109196	-1.760568
THYMINE	-1.53428705	-1.75176130	29.96011645	1.564052	0.131585	-2.008705
THYMINE	1.19931770	-1.72877957	33.36419682	2.271715	0.168005	-2.040369
THYMINE	3.40407717	0.09274343	36.66452153	2.846568	0.200428	-1.936760
THYMINE	4.06658674	2.97947326	39.87235027	-2.910786	0.218334	-1.782892
THYMINE	2.85712855	5.73255638	43.03872760	-2.402415	0.219068	-1.610790
THYMINE	0.27441268	7.20892763	46.23348818	-1.879513	0.199625	-1.454431
THYMINE	-2.58188696	6.83103172	49.51440636	-1.299392	0.166294	-1.356310

References

[1] Smith, T. ; Slipchenko, L. V. ; Gordon, M. S. In *Proceedings of 239th ACS National Meeting*, San Francisco, CA, 2010.

[2] Dabkowska, I. ; Jurečka, P. ; Hobza, P. *J. Chem. Phys.* **2005**, *122*, 204322.

[3] Ran, J. ; Hobza, P. *J. Phys. Chem. B* **2009**, *113*, 2933.

[4] Bravaya, K.B. ; Kostko, O. ; Ahmed, M. ; Krylov, A.I. *Phys. Chem. Chem. Phys.* **2010**, *12*, 2292.

[5] Chai, J.-D. ; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 084106.

[6] Jurecka, P. ; Sponer, J. ; Cerny, J. ; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.

[7] Sponer, J. ; Leszczynski, J. ; Hobza, P. *J Phys. Chem.* **1996**, *100*, 5590.

[8] Sponer, J. ; Leszczynski, J. ; Hobza, P. *Biopolymers* **2002**, *61*, 3.