THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

On the Role of London Dispersion Forces in Biomolecular Structure Determination

Michal Kolář,[†] Tomáš Kubař,[‡] and Pavel Hobza*[,†,§,∥]

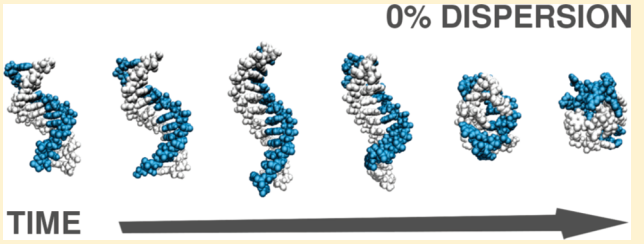
[†]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic

[‡]Institute of Physical Chemistry, Karlsruhe Institute of Technology, Kaiserstrasse 12, 76131 Karlsruhe, Germany

[§]Regional Center of Advanced Technologies and Materials, Department of Physical Chemistry, Palacký University, Olomouc, 771 46 Olomouc, Czech Republic

[∥]Department of Chemistry, Pohang University of Science and Technology, San 31, Hyojadong, Namgu, Pohang 790-784, Republic of Korea

ABSTRACT: A DNA dodecamer and the methyladenine···methylthymine (mA···mT) complex in aqueous environment have been studied by means of molecular dynamics simulation, with a modified force field accounting for the hypothetical absence of London dispersion forces. Under these conditions, the mA···mT complex is preserved, while the double-helical DNA oligomer passes via an extended, ladder-like intermediate to a collapsed structure. The results are interpreted in terms of stability and specificity of the structure of studied complexes. While the hydrophobic effect of the solvent accounts for the sufficient stabilization of the complex, the appearance of the native biomolecular conformation is attributed to the London dispersion forces. Thus, the London dispersion seems to provide the native structure of a biomolecular complex with the largest additional stabilization, preferring it among several (or many) possible aggregated structures. The observations are affected by the construction of the modified force field, and this effect is discussed thoroughly. The fundamental issues are the coupling of the components of the Lennard-Jones potential and the way to separate them. Based on the observations, the description of nonbonded interactions with the current biomolecular force fields is discussed. It is proposed that a novel force field composed of physically correct components to describe nonbonded interactions could exhibit more favorable performance in certain up-to-date applications.



1. INTRODUCTION

The structure of biomolecules such as proteins and nucleic acids is determined by noncovalent interactions, among which hydrogen bonding and the stacking interaction are prominent.[1] Their nature is quite different: hydrogen bonding arises from the electrostatic interaction between the polar proton donor and the proton acceptor as well as from the charge-transfer interaction of the electron-rich area of the proton acceptor and the σ^* orbital of the proton donor. On the other hand, the stacking interaction is almost exclusively conducted by the London dispersion forces, i.e., a genuinely quantum effect of the attraction of two fluctuating electron densities.[2] In an aqueous environment, there is yet another effect in action. Bulk water features a rich hydrogen-bonding network in the thermodynamic equilibrium, and this mesh has to be disrupted upon the insertion of a solute molecule. This perturbation costs a certain amount of (free) energy, and the tendency of water to minimize this cost is the hydrophobic effect.[3]

Different kinds of noncovalent forces may be involved in affecting the strength (structure stabilization) and specificity (structure determination) of biomolecular interactions. For instance, highly directional hydrogen bonds are claimed to ensure the recognition of complementary strands of DNA. The London dispersion has been mentioned in connection with the strength of the nucleobase stacking, biomolecular stabilization, protein folding and protein...ligand binding. In our previous studies, we have shown that the stabilization energy of nucleobase stacking reaches several tens of kJ mol^{-1} and plays an important role in the stabilization of the double-helical structure of DNA.[4] On the basis of experimental results, it has been suggested that the association of a ligand with a biomolecule can be attributed to the strong dispersion interaction between them.[5] These NMR and calorimetric studies of the binding of primary alcohols show the dominant effect of the dispersion interaction, and this conclusion is supported further by quantum-chemical calculations.[6,7] Also, the importance of dispersion interaction was highlighted by the results of an experiment on DNA modified with nonpolar nucleobase analogs: when strong hydrogen-bonded guanine−cytosine pairs in DNA were replaced by pairs of nonpolar nucleobase analogs without any hydrogen bonds, the stability of DNA (in terms of the melting point) increased surprisingly.[8] This means that the stacking of the base analogs governed by the dispersion interaction was able to compensate fully for the strong hydrogen bonding in the guanine−cytosine base pairs.

Received: March 28, 2011
Revised: May 13, 2011
Published: May 16, 2011

The early work in several laboratories starting in the 1960s tackled the issue of the relative magnitude of noncovalent interactions within such molecular systems.[9−12] The aggregation of hydrophobic molecules (e.g., hydrocarbons) owing to the action of the surrounding water was advocated by Honig et al.,[13,14] who pointed out that the dispersion attraction between hydrophobic systems is roughly equally strong as that between a hydrophobic system and water. An important concluding point was that "dispersion forces make no net contribution to the hydrophobic effect for liquid hydrocarbons".[14]

Modeling the dispersion interaction is rather difficult because of its quantum character. Accurate correlated ab initio calculations that account for dispersion energy properly are prohibitively demanding in terms of computer time and thus applicable only for molecular complexes of about tens of atoms.[15] On the other hand, the popular density-functional-theory methods, which are several orders of magnitude more efficient, fail completely to describe the mid-range correlation.[16−18] There have been attempts to account for the dispersion interaction by designing new functionals.[19,20] Eventually, the DFT methods were improved by the introduction of an empirical dispersion term.[21,22] However, they are still impractical for molecular dynamics simulations. The method of choice are still the empirical force field simulations.

Molecular dynamics (MD) simulation is a powerful tool to study biomolecules in their physiological environment.[23,24] These simulations are based on molecular mechanics force fields, which provide reasonably accurate results when used properly. The majority of currently used empirical force fields describe interactions between atoms by simple functions, which makes it possible to investigate the dynamics of very large molecular systems (of up to millions of atoms) on long time scales (of up to microseconds). Hence, computer simulation is very flexible in the modeling of physicochemical processes. Often, we can utilize a nonphysical simulation to explain a completely physical process. An illustrative example can be found in advanced free energy calculations involving alchemical mutations, where an atom may be for instance half hydrogen and half oxygen at the same time.[25−27] Encouraged by the good agreement of the dispersion energies obtained with empirical force fields with the exact calculations[28] and by the transparent character of the empirical force field, we decided to tackle the question of how important dispersion actually is for biomolecules, with a computer simulation using a modified empirical force field. Our approach is to simulate the molecular systems of interest with a potential that does not involve dispersion interaction. First attempts have already been made,[29,30] but as we show in the present study, a revision of the simulation protocol and deeper insight into the simulation characteristics are needed.

As in most of the molecular mechanics force fields, the nonbonded interaction energy in the Amber force field[31] is composed of two parts: the electrostatic and Lennard-Jones (LJ). The latter energy contribution contains an attractive (i.e., negative, dispersion) and a repulsive (i.e., positive, Pauli repulsion) term and can be expressed by two functional forms (eqs 1 and 2)

$$E_{\mathrm{LJ}} = \sum_{i<j} 4U_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (1)$$

$$E_{\mathrm{LJ}} = \sum_{i<j} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \quad (2)$$

which are equivalent for nonzero parameters A and B. r_{ij} is the interatomic distance, and U_{ij}, σ_{ij}, A_{ij}, and B_{ij} are Lennard-Jones pair parameters for which the following conversions hold (eqs 3 and 4)

$$A_{ij} = 4U_{ij}\sigma_{ij}^{12} \quad (3)$$

$$B_{ij} = 4U_{ij}\sigma_{ij}^{6} \quad (4)$$

Apparently, the repulsion and dispersion contributions can be separated conveniently using the latter functional form (eq 2).

Recently, we have examined the accordance of the attraction part $-B_{ij}/r_{ij}^{6}$ of the Lennard-Jones interaction energy in the Amber parm99 force field with the DFT-SAPT dispersion energy.[32−35] DFT-SAPT is an exact ab initio perturbation method to decompose the interaction energy of a complex into physically relevant contributions (e.g., electrostatic, exchange, dispersion, etc.). In the Amber family of force fields originally, the LJ parameters were adjusted to reproduce the macroscopic properties of the molecular system, such as the densities and vaporization enthalpies of simple organic liquids.[36,37] Despite the rather crude parametrization, we found that the attractive component of LJ potential approximates the reference DFT-SAPT dispersion energy quite well.[28,38] Also, it was found that the agreement may depend on the particular molecular complex geometry owing to the unphysical isotropy of the empirical LJ dispersion parameters and that the agreement of the force-field dispersion with the exact calculations may be better for molecules containing a wide range of atom types, which is indeed the case in the simulations of biomolecular systems. We note that the representation of the repulsion with a r^{-12} term is incorrect in principle as the correct dependence would be exponential; the algebraic form had been chosen primarily to improve computational efficiency and simplicity, since sufficient accuracy is guaranteed by a careful parametrization.[37,38]

The aim of this work is to assess the role London dispersion plays in the structure of biomolecular complexes. Previously, Černý et al.[29,30] performed simulations of a DNA system as well as of a small protein with a modified force field not accounting for the London dispersion, and observed huge conformational changes. They scaled the parameter U_{ij} in eq 1 by a factor of 0.01 to mimic the loss of dispersion. In the present work, we have extended the approach by Černý et al. with a careful design and a thorough discussion of several schemes to remove the dispersion interaction.

The absence of dispersion can be modeled in the simplest way by scaling down or even discarding the term $-B_{ij}/r_{ij}^{6}$ in the LJ functional form in eq 2 while keeping the term $+A_{ij}/r_{ij}^{12}$ intact. This is what we do for the interactions involving the solute molecule, along the entire range of interatomic distances, in our first scheme. However, there is an important side effect of such an intervention into the force field. The magnitude of the hydrophobic effect depends on the size of the solute;[3,39] that is, the disruption of the water structure is more pronounced with a larger solute. In fact, the situation is not straightforward since a possible favorable interaction of the solute with water molecules coparticipates in the overall hydrophobic effect of the surrounding water.

The repulsive component of the LJ force-field term comes into play as it controls the effective size of the atoms and molecules. More accurately, while the atom size in the molecular dynamics simulations results from a combination of the r^{-6} and r^{-12} terms in a standard force field, only the term r^{-12} matters in the modified force field stripped of the attractive component. Consequently, we will observe a certain increase of the effective size of the atoms and molecules once we have removed the dispersion in

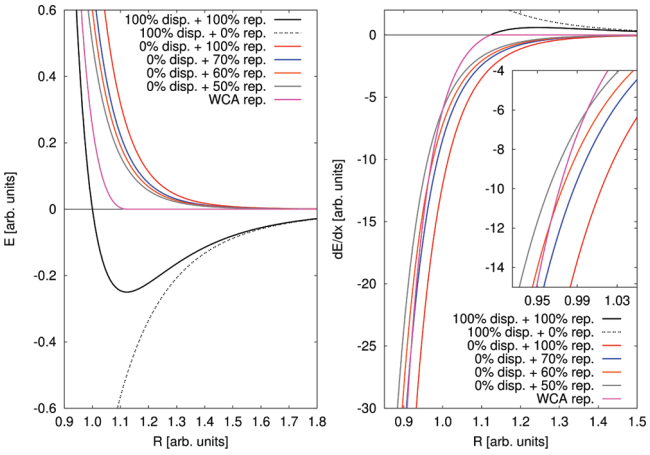


Figure 1. Lennard-Jones potential with its components and the scaled down repulsion and WCA term. Left, energy; right, gradient of energy.

the described way. Since this phenomenon may bias the results of the study to a large extent, it is desirable that another way to remove the London dispersion from the force field be found.

The other approach we adopt goes back to the work by Weeks, Chandler, and Andersen (WCA),[40,41] who suggested decomposing the LJ potential not to the individual algebraic terms in the equation but rather based on the intervals where the complete potential is repulsive or attractive. In this context, the repulsive component consists of a (shifted) full LJ potential up to the minimum and vanishes beyond that point. Conversely, the attractive component is nonconstant only beyond the minimum, where it is identical to the full LJ potential. The definition of the WCA potential is given in eqs 5−7.

$$E^{\mathrm{WCA}} = E^{\mathrm{WCA}}_{\mathrm{rep}} + E^{\mathrm{WCA}}_{\mathrm{attr}} \tag{5}$$

$$E^{\mathrm{WCA}}_{\mathrm{rep}} = \begin{cases} E_{\mathrm{LJ}} + U, & r \le r_0 \\ 0, & r > r_0 \end{cases} \tag{6}$$

$$E^{\mathrm{WCA}}_{\mathrm{attr}} = \begin{cases} -U, & r \le r_0 \\ E_{\mathrm{LJ}}, & r > r_0 \end{cases} \tag{7}$$

Here U is the energy depth of the LJ potential and $r_0 = (2)^{1/6}\sigma = (2A/B)^{1/6}$ is the position of the minimum. Obviously, the gradient of energy (cf. Figure 1) vanishes at r_0, is positive (attractive force) beyond r_0 and is negative (repulsive force) at distances shorter than r_0. Previously, the attractive component of WCA (eq 7) was used, e.g., to estimate the nonpolar (van der Waals) solvation free energy[42,43] and to study water interactions with carbon nanostructures.[44] In this study, we will make use of the repulsive WCA component (eq 6) consisting of the short-range part of LJ potential; this should guarantee a correct effective size of the atoms while contributing no attractive forces whatsoever.

The present work shall contribute to the assessment of the significance of the London dispersion for the biomolecular complexes. The studied cases are (1) the association free energy of the methyladenine···methylthymine complex in water and the effect of the loss of dispersion on the free energy, and (2) the structure and dynamics of a DNA oligomer in water simulated with a dispersion-free force field. The possible ways to improve the simulation protocols used so far will be discussed.

2. METHODS

2.1. Representation of Isolated Nucleobases. The geometries of the isolated methyladenine (mA) and methylthymine (mT) molecules were obtained with a B3LYP/cc-pVTZ[45−47] gradient optimization, and RESP atomic charges[48] were calculated at the HF/6-31G* level.[49,50] The Gaussian program was used[51] for all of the quantum-chemical calculations. The bonded and LJ parameters were assigned from the General Amber force field (GAFF),[52] which is suitable for a wide range of organic molecules. The geometry of the stacked mA···mT complex was prepared manually from the respective gas-phase geometries of the molecules.

The complex was solvated with ca. 2000 TIP3P water molecules[53] in a periodic cubic box with an edge length of ca. 4 nm. Electrostatic interaction was treated with the particle−mesh Ewald algorithm (PME)[54] with a direct-space cut off of 1.0 nm. A cutoff of 1.0 nm was used for the LJ interaction. The water molecules were minimized to avoid possible close contacts. A 20-ps NVT MD run was performed to heat the system to 300 K and was followed by a 1 ns NpT equilibration at 300 K and 1 bar. For heating and equilibration, the Berendsen thermostat and barostat were used, respectively.[55] Then, a harmonic potential with a force constant of 500 kJ mol^{-1} nm^{-2} and equilibrium distance of 0.35 nm was introduced and the solvated complex was equilibrated for another 0.5 ns at 300 K and 1 bar. All of the bonds were constrained to their reference lengths with the LINCS algorithm,[56] which made it possible to use the time step of 2 fs. The last snapshot of the equilibration was used as the starting structure for the umbrella sampling simulations.

2.2. Representation of the DNA Dodecamer. The double-helical DNA dodecamer of the d(CGCGAATTCGCG)$_2$ sequence was adopted from the crystal structure (PDB ID 1BNA).[57] The atom types and parameters were assigned from the Amber parm99 force field[31] with the correction for the backbone dihedral angles (parmBSC0).[58] The dodecamer was solvated with ca. 10,000 TIP3P water molecules in a periodic dodecahedron box. To achieve electroneutrality, 22 water molecules were replaced by Na^+ ions according to the electrostatic potential. All of the bonds were constrained to their reference length with the parallel version of LINCS.[56] For the electrostatics, PME was used with 1.0 nm direct-space cutoff, and 1.0 nm cutoff was used for the LJ potential.

The water molecules were the first to be minimized, followed by the minimization of the complete system. Then, the water was heated to 300 K while the DNA was restrained and its temperature held at 10 K. Subsequently, the entire system was heated to 300 K, followed by a 1 ns NpT equilibration at 300 K and 1 bar. The Berendsen thermostat and barostat were used for heating and equilibration, respectively.[55] The final structure was used as the initial structure for three 100 ns production simulations, which were performed with different initial velocities drawn randomly from a Maxwell−Boltzmann distribution at 300 K. For production simulations, the Nosé−Hoover thermostat[59,60] and the Parrinello−Rahman[61] barostat were used to sample the NpT ensemble. The coordinates were saved every 1 ps. All of the MD simulations were performed with the Gromacs 4 program package.[62]

2.3. DNA Anaylsis. For all of the trajectories, the time series of the following characteristics of the DNA dodecamer were evaluated: the radius of gyration, the root mean square deviation

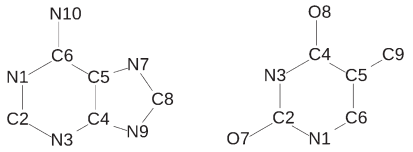


Figure 2. Numbering of the heavy atoms of methyladenine (left) and methylthymine (right).

(rmsd) and the solvent accessible surface area (SASA).[63] The rmsd was calculated for heavy atoms with respect to the first frame of the trajectory, which was identical for all of the simulations. The SASA was calculated with a probe radius of 0.15 nm. Tools from the Gromacs package as well as the VMD program[64] were used for the analysis.

2.4. Force Field Modifications. Two schemes were used to simulate the loss of dispersion interaction.

A. The term $-B_{ij}/r_{ij}^6$ was removed from the LJ functional form in eq 2. Since the Amber parm99 force field provides atomic U_i and σ_i LJ parameters, the A_{ij} and B_{ij} pair LJ parameters had to be obtained according to eqs 3 and 4, using also the combination rules[37,65]

$$U_{ij} = \sqrt{U_i U_j} \tag{8}$$

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j) \tag{9}$$

The parameter B_{ij} was set to zero for every interaction involving an atom of solute (i.e., mA, mT, or DNA). Also, a series of simulations were performed, where the repulsive term A_{ij}/r_{ij}^{12} was scaled down by a factor of 0.5, 0.6, and 0.7 to compensate for the undesirable increase of the effective atom size.

B. The attractive part of the LJ potential was removed with the WCA approach. Every pair interaction involving a solute atom was described merely with the repulsive WCA component (eq 6) instead of a full LJ interaction. The WCA repulsive potential was implemented in the Gromacs package.

The Lennard-Jones 1−4 interactions required for the correct description of dihedral angles in the Amber force field were conserved systematically.

2.5. Free Energy Calculations. The association free energy of the mA···mT complex was evaluated as the potential of the mean force with the umbrella sampling technique.[27,66] The distance between the C4 atom of mA and the C2 atom of mT (see Figure 2) was chosen as the reaction coordinate. We observed that the free energy profile does not depend on the choice of the atoms considered (not shown) as long as the atoms are located near the centers of mass of the molecules. A biasing harmonic potential was introduced between the carbon atoms (eq 10).

$$V_{\text{bias}} = \frac{1}{2}k(R - R_0)^2 \tag{10}$$

where k is the force constant, R_0 is the reference distance, and the variable R is the distance between mA(C4) and mT(C2). The force constant k was optimized to yield a smooth free energy profile of the association/dissociation process (not shown). The results presented below were obtained with a force constant of 1000 kJ mol^{-1} nm^{-2}, which seemed to be weak enough to allow a good sampling of the configuration space while strong enough to scan the reaction coordinate efficiently.

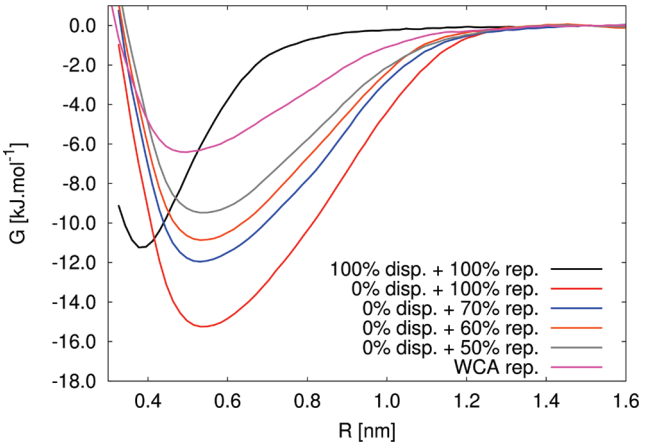


Figure 3. Association/dissociation free energy profile of the mA···mT complex in an aqueous environment, calculated with the unmodified (black line) as well as the various modified empirical force fields.

The reaction coordinate was split into 25 windows, with the reference distance of the biasing potential being varied between 0.30 and 1.50 nm with a step of 0.05 nm. Each window was simulated for 4 ns at 300 K and 1 bar, using the Nosé−Hoover temperature coupling and the Parrinello−Rahman pressure coupling. All of the bonds were constrained to their reference lengths (defined in the GAFF force field) with the LINCS algorithm, and the time step was 2 fs. The coordinates of atoms were saved every 200 fs. The starting structure of the first window (C4−C2 distance of 0.30 nm) was taken from the equilibration simulation, and the final structure from every simulation was taken as the starting structure for the simulation in the following window. The starting velocities were assigned from the Maxwell−Boltzmann distribution at 300 K, in each window. The initial interval of 1 ns in each window was considered as equilibration and was discarded before the analysis.

The probability density functions of the distance were evaluated and analyzed with the weighted histogram analysis method.[67] The uncertainty was estimated by comparing the free energy values obtained from the second, third and fourth nanoseconds of the windows. The described simulations were performed and the free energies were calculated with all of the force field modifications mentioned above as well as with the normal, unperturbed LJ potential.

2.6. Electrostatic Interaction Energy. An additional simulation of 200 ns was performed for the 0.40 nm window as well as for the 1.50 nm window. Here, the electrostatic (E_{elec}) component of the interaction energy between the nucleobases and the surrounding water was evaluated every 400 ps. The nonbonded cutoff was increased to 1.4 nm for the purpose of these calculations of energy.

3. RESULTS AND DISCUSSION

3.1. mA···mT Complex. The calculated profiles of the free energy vs distance are plotted in Figure 3. The plots are constructed from the data of the last 3 ns of each sampling window. The free energy features a single minimum, in agreement with the previous finding that the stacked configuration of complexes of aromatic species is highly preferred in an aqueous environment.[9] The values of the dissociation free energy obtained as the depth of the free energy minimum are summarized

Table 1. Dissociation Free Energy of the mA···mT Complex in an Aqueous Environment[a]

simulation	ΔG_{dissoc}
100% disp. + 100% rep.	11.2 ± 0.1
0% disp. + 100% rep.	15.3 ± 0.7
0% disp. + 70% rep.	12.0 ± 0.1
0% disp. + 60% rep.	11.0 ± 0.9
0% disp. + 50% rep.	9.5 ± 0.0
WCA rep.	6.4 ± 1.0

[a] The mean values and standard deviations of the values obtained from the second, third, and fourth nanosecond of the simulation windows are presented. All of the values are shown in kJ mol^{-1}.

in Table 1. The dissociation free energy of the mA···mT complex in an aqueous environment amounts to 11.2 kJ.mol^{-1} with the unmodified GAFF force field. It is not critical for the aims of the study to have determined this value exactly; yet, the results exhibit a good convergence as well as a reasonable qualitative agreement with previous results, both experimental and computational.[10,14,68−72]

The modification of the force field brings quite surprising results. Upon the removal of the dispersion interaction of the atoms of the solute ("0% disp. + 100% rep."), the position of the minimum shifts to longer distances by ca. 0.1 nm and the dissociation free energy increases to 15.3 kJ mol^{-1} (Table 1 and Figure 3). Apparently, the London dispersion is responsible for two competing kinds of interactions within the complex: First, there are attractive dispersive forces between mA and mT, which facilitate the formation of the complex, increasing the dissociation free energy. Second, there is an attractive dispersion interaction between each molecule (mA, mT) and the surrounding water molecules, which acts against the formation of the complex, as this interaction would be maximized for entirely separated mA and mT. The modified force field discards both of these antagonistic contributions at the same time, and there is little change to the free energy profile or the position of the minimum. This suggests that the contributions of the dispersive forces (1) between mA and mT and (2) between the molecules and the water are of comparable magnitude.

The dissociation free energy was further evaluated with the repulsive term of the LJ potential scaled down (Table 1). A clear trend was observed: lower dissociation energy is obtained with stronger damping of the repulsion. With the dispersion removed and repulsion scaled to 50%, the dissociation free energy amounted to 9.5 kJ mol^{-1}.

However, the modification of the force field brings on a spurious change of the structure of the system: the distribution of the water molecules around the solute is different from that obtained with the full LJ potential (see the illustrative radial distribution function presented in Figure 4). The water penetrates slightly closer to the solute atoms, but the amount of water present (expressed as the height of the first peak) is much smaller than that which is obtained with the full potential. This observation correlates with the shape of the modified energy function and its gradient vs distance: the gradient is slightly less negative on short distances but increases less steeply compared to the unmodified case (Figure 1, right, inset). It is possible to say that atoms are softer and larger with the modified potential. In turn, the (albeit tiny) change of the interatomic distances affects for instance the strongly distance-dependent electrostatic interactions. This is

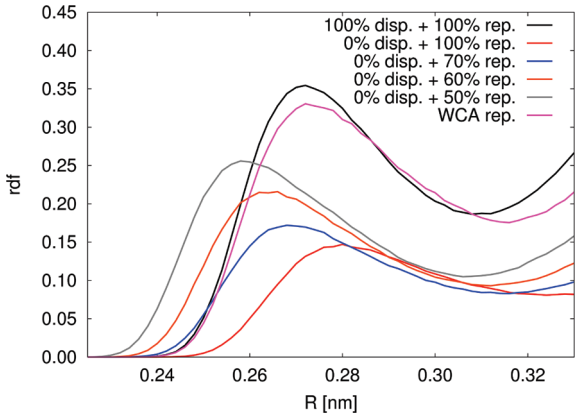


Figure 4. Radial distribution function of the water oxygen atoms around the heavy atoms of the solute, calculated from the simulations of DNA with the various force field modifications.

Table 2. Sum of All of the Pair Electrostatic Interactions between the Nucleobase Molecules (mA, mT) and a Water Molecule[a]

simulation	E_{el} (0.4 nm)	E_{el} (1.5 nm)	$\Delta E_{el,dissoc}$
100% disp. + 100% rep.	−231	−248	−17
0% disp. + 100% rep.	−131	−151	−20
0% disp. + 70% rep.	−175	−197	−22
0% disp. + 60% rep.	−199	−223	−24
0% disp. + 50% rep.	−232	−257	−25
WCA rep.	−217	−241	−24

[a] The energies were calculated every 400 ps and averaged over the simulation of 200 ns. All of the values are shown in kJ mol^{-1}. The equilibrium distance of the biasing umbrella potential is provided in the parentheses.

obviously undesirable as the thermodynamics of the system is perturbed in a way that is difficult to characterize.

The described problem is caused by the unsuitable modification of the LJ potential; the balance between the terms $+A_{ij}/r^{12}$ and $-B_{ij}/r^{6}$ seems to be indispensable for keeping the structure of the system. This issue cannot be resolved by a simple scaling of the repulsion term. For this reason, we decided to employ the WCA decomposition scheme, which keeps both terms for the short-range interaction ($r < r_0$). Indeed, the radial distribution function of water around the solute (Figure 4) changed negligibly with the WCA repulsion potential as compared to the full LJ potential. The dissociation free energy with WCA repulsion was lowered to 6.4 kJ mol^{-1}, and the optimal distance (minimum of free energy) was longer, ca. 0.48 nm.

Also, the electrostatic interaction energy between mA···mT and the surrounding water was evaluated both for the molecules in the complex and for the molecules separated, see Table 2. The difference of these quantities is a component of the dissociation enthalpy, with negative values corresponding to a stronger attractive interaction in a separated state, thus favoring a dissociation of mA···mT. The loss of dispersion affects the absolute magnitude of the electrostatic interaction enormously (by as much as 100 kJ mol^{-1}), but the effect on the change upon dissociation is much smaller. By also scaling down the repulsion term, the deviation of the electrostatic interaction energy from the unmodified case becomes more pronounced: the ΔE_{elec}

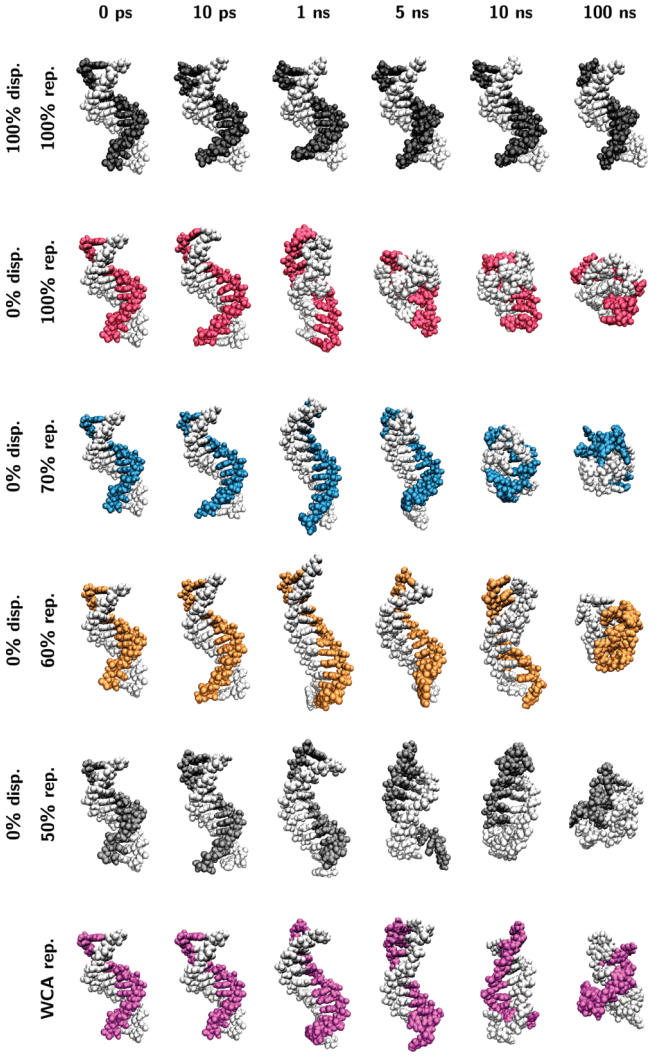


Figure 5. Representative structures of the DNA oligomer in simulations performed with the various modified force fields. The colors correspond to the lines in the other figures.

upon dissociation is -25 kJ mol^{-1} with the dispersion removed and repulsion scaled to 50%, as compared to -17 kJ mol^{-1} with the full LJ potential. Evidently, the change of the form of a component of the MM energy function has an influence on other kinds of interactions as well. This effect proceeds indirectly via the structure of the molecular system, and demonstrates that great care has to be taken when handling the modified empirical force field.

3.2. DNA Dodecamer. Three MD simulations of 100 ns each were performed for a DNA dodecamer in an aqueous solution with the standard Amber force field as well as with each of the modified force fields described above. The set of the representative structures from the simulations is presented in Figure 5. The time evolution of the radius of gyration, of the rmsd and of the SASA are shown in Figures 6−8, respectively. Note that the starting structure was identical for all of the simulations, and only the initial velocities varied. The simulation with the full LJ potential (the black lines in Figures 6−8) shows stable behavior considering all of the monitored characteristics.

The starting structure in all of the MD simulations was in the double helical conformation that resulted from the equilibration

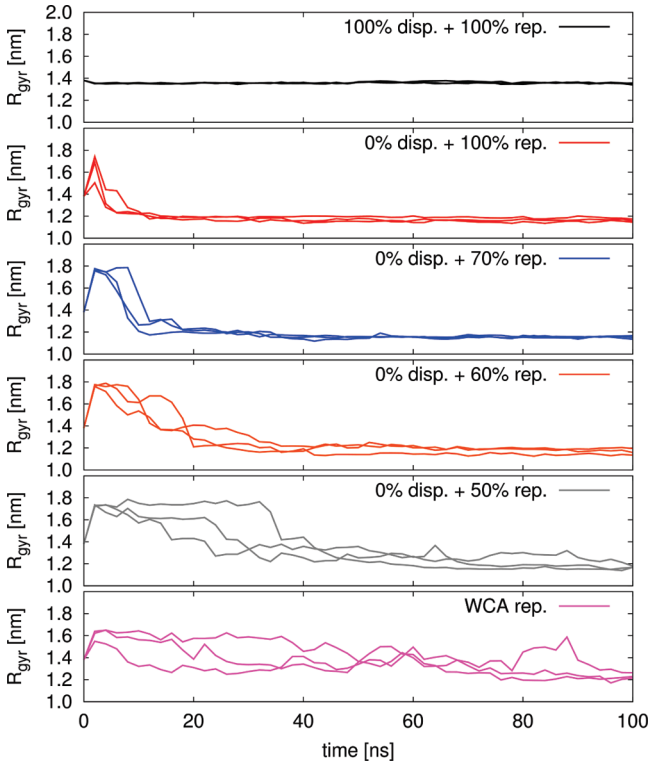


Figure 6. Time evolution of the radius of gyration of the DNA oligomer. The values were collected every picosecond, and the averages over the windows of 1 ns are plotted.

with the full LJ potential. Thus, the simulations with modified potentials were far from equilibrium initially, and a relaxation was to be expected in the simulation. Indeed, large structural rearrangement was observed upon canceling the London dispersion ("0% disp. + 100% rep."): The DNA elongates within the first nanosecond, and the distance between the base-pair steps increases. The radius of gyration increases steeply from the starting value of 1.38 nm. Then, within a few tens of nanoseconds, the hydrogen bonds between the nucleobases are cleaved, and the DNA collapses into a very compact structure. The process is depicted in Figure 5 and is reflected by a gradual decrease of the radius of gyration to ca. 1.2 nm, a value that is ca. 0.2 nm smaller than that of the double helical structure (see Figure 5, the red structures, and Figure 6, the red lines).

Applying the same scaling of the LJ repulsion term as in the case of the mA···mT complex, we obtained a series of trajectories with different effective size and softness of the DNA atoms. Generally, it can be said that the amount of repulsion conserved determines how quickly the system equilibrates. The reason seems to be in the first step of the process, the cleavage of hydrogen bonds. In a simulation with the strongest repulsion ("0% disp. + 100% rep."), the hydrogen-bonded atom pair is prevented from maintaining the short-ranged arrangement more strongly, and the DNA strands tear apart in a shorter time than in a simulation with scaled repulsion. The subsequent collapse is reflected by a smooth decrease of SASA (shown in Figure 8).

In the simulations with the WCA repulsion, there is little undesired perturbation of the hydrogen bonding patterns to be expected. Still, the DNA relaxes through an extended structure into a collapsed one as well. The observed fluctuations of the radius of gyration, rmsd as well as SASA are larger than in the

dx.doi.org/10.1021/jp202878d |J. Phys. Chem. B 2011, 115, 8038–8046

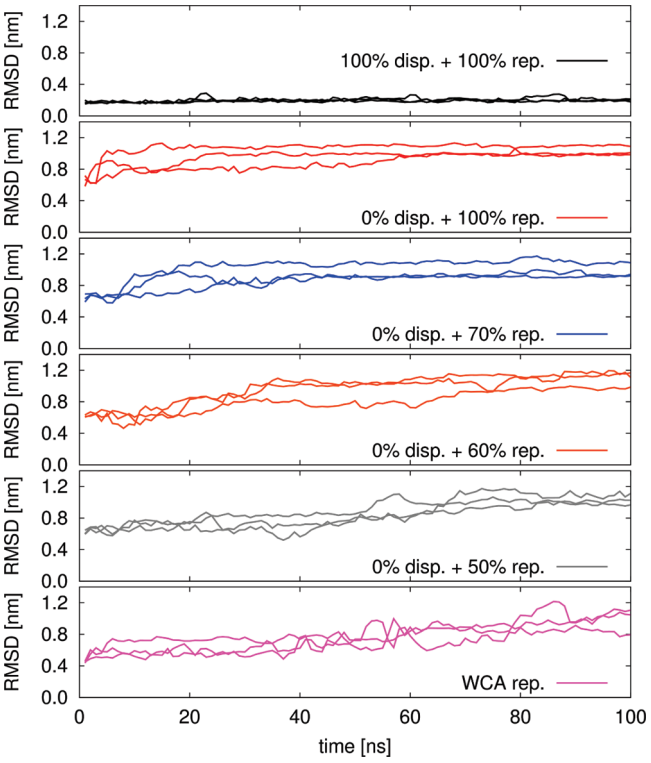


Figure 7. Time evolution of the root-mean-square deviation (rmsd) of the heavy atoms of DNA from the starting structure, which corresponds to the double-helical conformation. The values were collected every picosecond, and the averages over the windows of 1 ns are plotted.

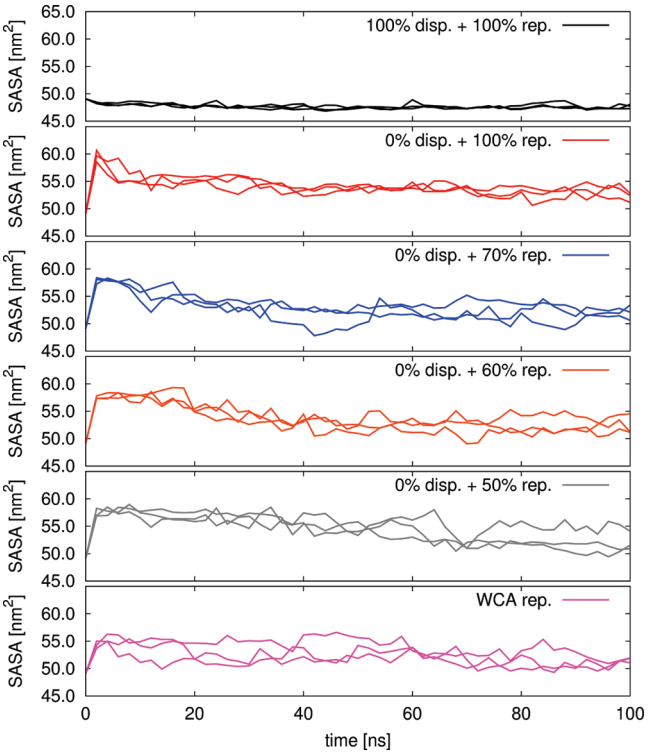


Figure 8. Time evolution of the solvent accessible surface area (SASA) calculated with a probe radius of 0.15 nm. The individual lines in each panel stand for the trajectories with various initial velocities. The values were collected every picosecond, and the averages over the windows of 2 ns are plotted.

simulations with the other force field modifications (Figures 6−8, the magenta lines). Eventually, the DNA strands do not separate upon the removal of the London dispersion. Thus, the action of dispersive (or generally attractive) forces seems to be responsible for the determination of particular structural features rather than for the stabilization of an aggregate in aqueous environment.

The intermediate extended structure with a large radius of gyration observed in our simulation resembles closely the ladder-like DNA structure proposed on the basis of pulling simulations.[73] Interestingly, according to the report by Albrecht et al.,[74] this structure is likely to occur only if the stretching is sufficiently fast, as is always the case in MD simulations with their loading rates being several orders of magnitude larger than those in experiments. This points out to the metastable character of the extended structure, even though the perturbation of DNA structure is brought about by an entirely different means (mechanical stress rather than a modification of the force field).

3.3. Role of Dispersion Interaction. The DNA strands do not dissociate in the simulations with modified force field, which is in correlation with the presence of a minimum on the free energy profile of the mA···mT complex lacking the dispersion interaction. The surrounding water acts to minimize its own structural perturbation, which has been caused by the extension of the DNA double helix. However, the double-helical structure of DNA does not recover (in spite of its similar spatial extension), and the DNA assumes a disordered collapsed conformation instead. It is worth emphasizing the nonspecificity of the hydrophobic effect demonstrated by the fluctuation of rmsd (Figure 7), which is much larger in the case of DNA without dispersion than in the unmodified case. The structural variance of the DNA species in the collapsed state is richer than in the double-helical conformation undoubtedly. The same effect is observed for the mA···mT complex, where the minimum of the association free energy is considerably wider when dispersion interaction is absent.

The surrounding water molecules guarantee the association of the solute, but lacking the dispersion interaction there are many solute structures which disrupt bulk water similarly. The significance of London dispersion for biomolecular complexes may be to ensure the "correct" structure/conformation of the complex in aqueous environment by providing an additional stabilization to one of the possible aggregates. The conformational space of the solute is restricted by the dispersion interaction significantly.

The structural fluctuations are temperature dependent, naturally. When the temperature decreases, it is reasonable to expect an increased stability of the DNA double helix. So, it would be of interest to perform simulations with a modified potential at a lower temperature, to investigate the balance between the forces keeping double-helical conformation and the disruptive effect of thermal fluctuations.

3.4. Description of Dispersion Interaction with Empirical Force Fields. The difference between the results by Černý et al.[29] and our work is that the extended structure (called "ladder-like" in ref 29) is not stable in this work, but rather an intermediate which collapses into a more compact structure subsequently. This is most likely caused by the scheme to remove the London dispersion applied by Černý et al.: The simple scaling down of the parameter U (see eq 1) to as little as 1% affected not only the attractive component of LJ potential (describing the dispersion

dx.doi.org/10.1021/jp202878d |J. Phys. Chem. B 2011, 115, 8038–8046

interaction) but also the repulsive one. Then, the too weak repulsive interaction allowed hydrogen-bonded atoms to approach closer to each other, making the hydrogen bonds artificially stronger because of stronger Coulomb interaction on the shorter distance. Such strong hydrogen bonds would not break and the DNA would remain in the ladder-like structure. However, this is not observed here if the repulsive potential is kept intact, as is the case with the WCA scheme. Here, the hydrogen bonding is not strong enough to support the ladder-like structure (with a large SASA). This structure collapses owing to the hydrophobic effect of the surrounding water. Apparently, the hydrophobic effect of water overcomes the hydrogen bonding within DNA, which is the major stabilizing interaction remaining after the removal of dispersion in our simulations.

The crucial task in this work was to find a way to remove the London dispersion from the force field with as little side effects as possible. While this issue seems to have been resolved by the choice of the WCA scheme to an appreciable extent, a more general point of ambiguity within current biomolecular force fields becomes obvious: Their functional form is different than that of a force field to be parametrized to well-defined components of interaction energy. Specifically, the London dispersion and the Pauli repulsion would be parametrized independently in such a force field rather than in a common expression like the LJ potential. The description with physically correct functions would be particularly beneficial in applications where the force field is required to perform outside of the region for which it was parametrized; the performance of current biomolecular force fields relies on the compensation of errors to a considerable extent[38] and may turn out to be insufficient in these cases. An example of such an application may be the broad field of simulation of biomolecular structures under mechanical stress, like extended structures of DNA and proteins.

Concentrating on the dispersion interaction, it has to be pointed out that the $-B/r^6$ term in current force fields penetrates to shorter interatomic distances where the dispersion interaction itself is ill-defined. Provided the Pauli repulsion is represented by a physically correct exponential function, appropriate parametrization of dispersion seems to be only possible if special care is taken on shorter interatomic distances. In this context, it is instructive to note the approximative DFT method SCC-DFTB,[75] which is completely free of the mid-range correlation energy. Here, an empirical correction for dispersion energy consisting of a damped $-B/r^6$ term was introduced with success,[76] and this may be the way to construct the dispersion component of a new-generation force field.

4. CONCLUSIONS

We have studied two biomolecular systems by means of MD simulations: the mA···mT complex in water and a double-helical DNA oligomer in water. The hypothetical condition of the loss of London dispersion was modeled by the removal of the attractive part of the Lennard-Jones interatomic potential in two different ways. In both of the studied systems (mA···mT and DNA), the aggregation of the solute was preserved even in the absence of dispersion forces. The mA···mT dissociation free energy did not change significantly, and the DNA strands remained aggregated upon the considered modifications of the force field. The structure of the DNA oligomer passed from the initial helical double-stranded through an extended double-stranded into a collapsed one, with all of the modified potentials.

Based on the observations, it is suggested that the hydrophobic effect of the surrounding water is probably strong enough to support an aggregated DNA structure though. However, the effect is nonspecific, unable to discriminate between structures of similar spatial extension (e.g., the double-helical and collapsed one). Instead, it is the dispersion forces that ensures the correct "recognition" of the structure in water: the double helix in the case of DNA.

The spurious side effects of the removal of dispersion were identified. The coupling between the attractive and the repulsive terms of the LJ potential affects the effective size as well as the softness of atoms in the simulation, compromising the fragile balance of Lennard-Jones and electrostatic energy contributions. This issue is alleviated practically by the implementation of the WCA modification, which leaves the size and the softness of atoms untouched.

The results of this study indicate that it might be meaningful to improve the design of the nonbonded part of the current biomolecular force fields. The most promising way is to parametrize the force field to the physically defined components of the interaction energy, with correct asymptotic behavior. Additionally, it would be necessary to provide the dispersion contribution with a damping function, to ensure a correct description of dispersion interaction on shorter interatomic distances where the dispersion is not defined unambiguously. Such a force field would be more likely than the current force fields to work well outside of the regions for which it was parametrized.

AUTHOR INFORMATION

Corresponding Author

*Tel.: (+420) 220 410311. E-mail: pavel.hobza@uochb.cas.cz.

ACKNOWLEDGMENT

The authors are grateful to Dr. Petr Jurečka for stimulating discussions and to the anonymous reviewer for valuable suggestions. This work was supported by the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic [Z40550506], the Ministry of Education, Youth and Sports of the Czech Republic [LC512 and MSM6198959216], the Czech Science Foundation [P208/11/0295] and Korea Science and Engineering Foundation [World Class Univ. program: R32-2008-000-10180-0]. This work was also supported by the Operational Program Research and Development for Innovations—European Science Fund (CZ.1.05/2.1.00/03.0058). The support of Praemium Academiae, Academy of Sciences of the Czech Republic, awarded to P.H. in 2007 is also acknowledged.

REFERENCES

(1) Hobza, P.; Müller-Dethlefs, K. *Non-covalent Interactions: Theory and Experiment*; RCS Publishing: London, 2009.
(2) London, F. *Z. Physik.* **1930**, *63*, 245–279.
(3) Chandler, D. *Nature* **2005**, *437*, 640–647.
(4) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247–3276.
(5) Malham, R.; Johnstone, S.; Bingham, R. J.; Barratt, E.; Phillips, S. E. V.; Laughton, C. A.; Homans, S. W. *J. Am. Chem. Soc.* **2008**, *127*, 17061–17067.
(6) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613.
(7) Vondrášek, J.; Bendová, L.; Klusák, V.; Hobza, P. *J. Am. Chem. Soc.* **2005**, *127*, 2615–2619.
(8) Řeha, D.; Hocek, M.; Hobza, P. *Chem.—Eur. J.* **2006**, *12*, 3587–3595.
(9) Sinanoglu, O.; Abdulnur, S. *Photochem. Photobiol.* **1964**, *3*, 333–342.

(10) Sowers, L. C.; Shaw, B. R.; Sedwick, W. D. *Biochem. Biophys. Res. Commun.* **1987**, *148*, 790–794.
(11) Privalov, P. L.; Gill, S. J. *Adv. Protein Chem.* **1988**, *39*, 191–234.
(12) Creighton, T. E. *Curr. Opin. Struct. Biol.* **1991**, *1*, 5–16.
(13) Nicholls, A.; Sharp, K. A.; Honig, B. *Prot. Struc. Func. Gen.* **1991**, *11*, 281–296.
(14) Friedman, R. A.; Honig, B. *Biophys. J.* **1995**, *69*, 1528–1535.
(15) Riley, K. E.; Pitoňák, M.; Jurečka, P.; Hobza, P. *Chem. Rev.* **2010**, *110*, 5023–5063.
(16) Kristyán, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175–180.
(17) Perez-Jorda, J. *Chem. Phys. Lett.* **1995**, *233*, 134–137.
(18) Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 2365–2376.
(19) Andersson, Y.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **1996**, *76*, 102–105.
(20) Zhao, Y.; Truhlar, D. G. *Teor. Chem. Acc.* **2008**, *120*, 215–241.
(21) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473.
(22) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. *J. Comput. Chem.* **2007**, *28*, 555–569.
(23) Cheatham, T. E.; Young, M. A. *Biopolymers* **2000**, *56*, 232–256.
(24) Ponder, J. W.; Case, D. A. Protein Simulations. *Advances in Protein Chemistry*; Academic Press Inc.: San Diego, CA, 2003; Vol. 66, pp 27−85.
(25) Gilson, M. K.; Zhou, H. X. *Ann. Rev. Biophys. Biomol. Struc.* **2007**, *36*, 21–42.
(26) Shirts, M.; Mobley, D.; Chodera, J. *Ann. Rep.Comput. Chem.* **2007**, *3*, 41–59.
(27) Chipot, C.; Pohorille, A. *Free Energy Calculations*; Springer-Verlag: Berlin, 2007.
(28) Kolář, M.; Berka, K.; Jurečka, P.; Hobza, P. *ChemPhysChem* **2010**, *11*, 2399–2408.
(29) Černý, J.; Kabeláč, M.; Hobza, P. *J. Am. Chem. Soc.* **2008**, *130*, 16055–16059.
(30) Černý, J.; Vondrášek, J.; Hobza, P. *J. Phys. Chem. B* **2009**, *113*, 5657–5660.
(31) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struc. Dyn.* **1999**, *16*, 845–862.
(32) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887–1930.
(33) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *357*, 464–470.
(34) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *362*, 319–325.
(35) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2003**, *367*, 778–784.
(36) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
(37) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K . M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(38) Zgarbová, M.; Otyepka, M.; Šponer, J.; Hobza, P.; Jurečka, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10476–10493.
(39) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570–4577.
(40) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237–5247.
(41) Chandler, D.; Weeks, J. D.; Andersen, H. C. *Science* **1983**, *220*, 787–794.
(42) Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. *J. Am. Chem. Soc.* **2003**, *125*, 9523–9530.
(43) Shivakumar, D.; Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 919–930.
(44) Athawale, M. V.; Jamadagni, S. N.; Garde, S. *J. Chem. Phys.* **2009**, *131*, 115102–115109.
(45) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
(46) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(47) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
(48) Bayly, Ch. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
(49) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69–89.
(50) Hariharan, P. C.; Pople, J. A. *Theor. Chem. Acc.* **1973**, *28*, 213–22.
(51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
(52) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
(53) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(54) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(55) Berendsen, H. C. J.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
(56) Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.
(57) Dickerson, R. E.; Drew, H. R. *J. Mol. Biol.* **1981**, *149*, 761–786.
(58) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.
(59) Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.
(60) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
(61) Parrinello, M.; Rahman, A. *J. App. Phys.* **1981**, *52*, 7182–7190.
(62) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
(63) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379–400.
(64) Humphrey, W.; Dalke, A.; Schulten, K. *J. Molec. Graphics* **1996**, *14*, 33–38.
(65) Halgren, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 7827–7843.
(66) Torrie, G.; Valleau, J. *J. Comput. Phys.* **1977**, *23*, 187–199.
(67) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
(68) Guckian, K. M.; Schweitzer, B. A.; Ren, R. X. F.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 2213–2222.
(69) Nakano, N. I.; Igarashit, S. J. *Biochemistry* **1970**, *9*, 577–583.
(70) Cieplak, P.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 3734–3739.
(71) Dang, L. X.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 503–507.
(72) Stofer, E.; Chipot, C.; Lavery, R. *J. Am. Chem. Soc.* **1999**, *121*, 9503–9508.
(73) Lebrun, A.; Lavery, R. *Nucleic Acids Res.* **1996**, *24*, 2260–2267.
(74) Albrecht, C. H.; Neuert, G.; Lugmaier, R. A.; Gaub, H. E. *Biophys. J.* **2008**, *94*, 4766–4774.
(75) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, Th.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.
(76) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149–5155.