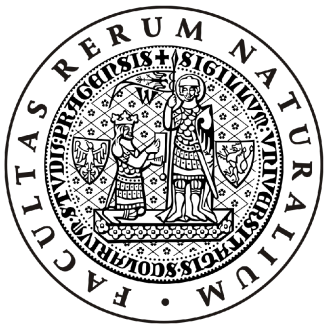
Univerzita Karlova v Praze, Přírodovědecká fakulta
Katedra fyzikální a makromolekulární chemie



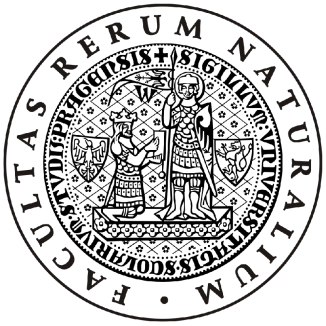
Diplomová práce

Úloha Londonových disperzních sil ve stabilitě struktury DNA: in silico experiment

Michal KOLÁŘ

Školitel: prof. Ing. Pavel Hobza, DrSc.
Akademie věd České republiky, Ústav organické chemie a biochemie
Centrum biomolekul a komplexních molekulových systémů
2009

Charles University in Prague, Faculty of Science
Department of Physical and Macromolecular Chemistry



Diploma thesis

The Role of London Dispersion Forces in Stability of DNA: in silico experiment

Michal KOLÁŘ

Advisor: prof. Ing. Pavel Hobza, DrSc.
Academy of Sciences of the Czech Republic,
Institute of Organic Chemistry and Biochemistry
Center for Biomolecules and Complex Molecular Systems
2009

Čestně prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně, pouze za použití citované literatury. Dále prohlašuji, že v této diplomové práci prezentuji výsledky své práce a nebo práce, na níž jsem se podílel.

Tato diplomová práce ani žádná její část nebyly dříve použity k získání jakéhokoliv akademického titulu.

V Praze, 27. dubna 2009

Michal Kolář

I would like to thank Prof. Pavel Hobza for his patient guidance during my research about the properties of DNA, for the stories he said to lighten the mood and for his understanding during the tough situations I had to face.

I thank Tomáš, who helps me not only in the scientific field. I also thank the students from Canon, because the working environment which they make every day *is* extraordinarily pleasant and I can not imagine better place to work (maybe except the Jeseníky mountains).

Many thanks pertains to my sister and parents. I know, it was as big task for them as it was for me.

I wish I could fly
Shirley Manson

Contents

List of Figures

List of Tables

List of Abbreviations

A-DNA	A-conformation of deoxyribonucleic acid
B-DNA	B-conformation of deoxyribonucleic acid
B3LYP	Becke exchange-correlation functional
cc-pVTZ	correlation-consistent valence triple zeta bases set
DFT	density functional theory
DNA	deoxyribonucleic acid
HF	Hartree-Fock
LINCS	linear constraint solver
LJ	Lennard-Jones
mA	9-methyladenine
MD	molecular dynamics
mT	1-methylthymine
PDB	protein database
PME	particle-mesh Ewald
pmf	potential of mean force
RESP	restricted electrostatic potential
SS	solute-solute
WHAM	weighted histogram analysis method
WS	water–solute
WW	water–water

Preface

When James Watson and Francis Crick discovered in 1953 the structure of deoxyribonucleic acid[1] , they opened the door into the new world of science. Since the fifties it has been believed, that the DNA might be the carrier of the genetic information. After the first publication, Watson and Crick proposed the possible mechanism of DNA duplication and consequently the mechanism of passing the genetic information from the mother cell to its descendants.[2]

For the man it is a very attractive idea to know the complete genetic code, to understand in which processes the DNA is involved and later maybe to comprehend why we are (or are not?) so special in nature. To catch up all the information we should start with precise understanding of the DNA structure which can hopefully lead us further. It has already been done a lot in the field of structural biochemistry and molecular biology, beginning with description of DNA components[3] ending with e.g. advanced experiments,[4] and the DNA became a topic well known even among grammar school pupils.

There are certainly many effects, which form DNA into double-helical conformation and play some role in the dynamics of DNA. The nature evidently cannot use the power of the covalent bonding in the living organisms in those situations, which need maintaining a delicate equilibrium between bound and unbound states. It is not surprising, that non-covalent interactions are in the sun. Besides hydrogen bonds between complementary bases which are believed being responsible for the faultless copying of the strands, the London dispersion forces have been awarded by an enhanced importance.[5,6] In this thesis we use the latest tools of computational chemistry to perform so called *in silico* experiment and try to answer at first glance a very simple question: *What would happen if we made DNA away with London dispersion forces?*

Chapter 1

Methods

1.1 Introduction

Computational chemistry have noted an extensive development in several past years and as a helping hand for this development significant progress in the computer industry has been serving. Nowadays, the molecular modeling is an equivalent partner of the experiments. The evidence for this might be *e.g.* the increasing number of the scientific journals concerning theoretical chemistry [7] or the Nobel prize in chemistry in 1998 for John A. Pople and Walter Kohn. [8] Even thought both of these gentlemen were awarded for their contribution to the non-empirical (or *ab initio*) calculations, one should bear in mind that for biologically relevant studies the size of the models goes to the hardly gainable altitudes for *ab initio* calculations and studies based on empirical potential embody a great success. [9–11]

The relevance of such models is often discussed and the agreement with an experimental data is crucial for the model being trustful. However, in the world of ones and zeros we are, strictly speaking, not restricted by any physical laws and our model can suffer from situations, which cannot occur in the real world. This, so called *in silico experiments*, which are more or less based on the physical reality, can introduce new information about studying system which is for experiment unreachable, but one should be still aware of the limitations of such model and the conclusions should be declare with care.

Indeed, reducing the dispersion forces belongs to such a sort of a *non-physical* simulations (apart from the other problems relating with the relevance of the *physical*

simulations) because dispersion forces are universal and there are *not* any atoms or molecules which would naturally lack dispersion forces. Nevertheless there are certain experimental manipulations which can be used to modify dispersion interaction, such as C=O→C=S substitution, the modification proposed bellow is unique and in principle experimentally inimitable. The author is honestly aware of that and all conclusion are delivered with the highest carefulness.

1.2 Simulations

Since we are interested particularly in dynamical behavior of DNA and in changes connected with the lack of dispersion forces, we need computational method which is able to describe time evolution of the system of interest. Moreover, since the size of the system under study is out of range of *ab initio* methods (including DFT), from which the cheapest are suitable for systems with less than approx. 1000 atoms, we are gently enforced to use molecular dynamics methods based on empirical potential.

1.2.1 Molecular Dynamics

Usually meaning molecular dynamics (MD) we describe the system classically, that means the time evolution of the system is covered by the integration of the classical Newton's equations of motion.

$$F = m\frac{\partial^2 x}{\partial t^2}. \tag{1.1}$$

Integration of Eqn. 1.1 is executed in the most simple case by Verlet algorithm.[12] For our calculations a modification of Verlet algorithm was used. The leap-frog[13] method, which surpasses the Verlet by numerical precision, updates positions and velocities in different moments.

$$r(t+\Delta t) = r(t) + v(t+\frac{\Delta t}{2})\Delta t \tag{1.2}$$

$$v(t+\frac{\Delta t}{2}) = v(t-\frac{\Delta t}{2}) + \frac{F(t)}{m}\Delta t \tag{1.3}$$

where t stands for the actual time, Δt is the time step, $v(t)$ stands for the velocity at time t, $r(t)$ stands for position at time t and $F(t)$ is the force acting at time t. In such integration the time step is an important parameter. In general it is claimed, that 5 numerical steps should be performed during one harmonic oscillator step[14] and the time step length is therefore limited by the fastest oscillations in the system. These are usually hydrogen oscillations. To deal with this, constrained bond lengths are introduced. Constraining the bonds allows us to use time step up s long as 2 fs and it also seems that constrained bonds approximate *quantum nature of the bonds* better than harmonic oscillator.[14] In all simulations we used LINCS constraint algorithm[15] to constraint all bonds which works efficiently also in parallel.[16]

The molecular dynamics simulations of the solvated systems are often performed under periodic boundary conditions. These allow us to cope with the fact that we are not able to simulate systems of macroscopic dimensions with atomistic resolution when atomistic resolution is needed and that the aperiodic system would suffer from a disnature boundaries of the solvated model and vacuum around. In this study, periodic boundary conditions were adopted with a rhombic dodecahedron box,[14] see Fig. 1.1.

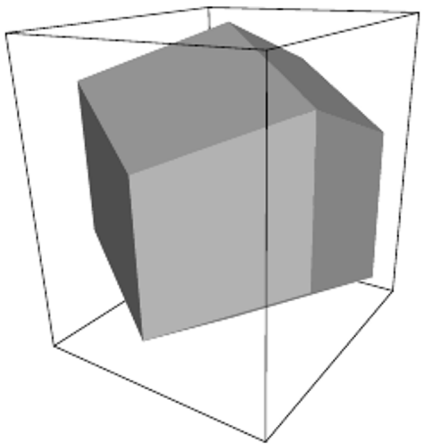


Figure 1.1: Rhombic dodecahedron box vs. cubic box. Adopted from Ref.[14]

The simulation box was exposed to a temperature bath, represented by an extended-ensemble Nosé-Hoover thermostat.[17,18] During equilibration and heating we used weak coupling Berendsen thermostat,[19] which is able to relax to the target tempera-

ture in much shorter time but compared to Nosé-Hoover scheme is not able to sample canonical ensemble correctly. All the simulations were carried out under temperature 300 K.

Because the chemical processes usually occur under a constant pressure, we used also pressure coupling, during equilibration Berendsen scheme[20] followed by the one proposed by Parrinello and Rahman[21] during production runs.

1.2.2 Interaction Potential

The crucial part of the molecular dynamics simulations are interaction potentials, which are responsible for the glorious success or misery and perdition of MD. The articles bringing widely used potentials like Amber, CHARMM or OPLS belong to the most cited ones[22–24] which confirms the large popularity of the calculations based on the empirical potential.

For the numerical integration according to Eqn. 1.3 we need to know the forces acting on all atoms. These are obtained as a first derivative of the potential energy, which is know in analytical form. In this study we modeled nucleic acid bases and DNA double-helix by Amber force field for which the total potential energy depends on the positions of all atoms ($\vec{x}$) and can be expressed as follows:

$$
\begin{aligned}
E_{tot}(\vec{x}) &= \sum_{bonds} K_r(r - r_{eq})^2 + && (1.4)\\
&+ \sum_{angles} K_\vartheta(\vartheta - \vartheta_{eq})^2 + && (1.5)\\
&+ \sum_{dihedrals} \frac{V_n}{2}\left(1 + cos(n\phi - \gamma_0)\right) + && (1.6)\\
&+ \sum_{i<j} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right) + && (1.7)\\
&+ \sum_{i<j} \frac{q_i q_j}{r_{ij}}. && (1.8)
\end{aligned}
$$

In the formula the term 1.4 describes bonds stretching, the term 1.5 angles bending, the term 1.6 stands for dihedral angles' torsions, the terms 1.7 and 1.8 cover

non-bonded interactions described by Lennard-Jones potential and Coulomb law, respectively. From the last two sums (1.7, 1.8) those interactions between the atoms connected via less than 3 bonds are excluded while non-bonded interaction between the atoms connected via 3 bonds (so called 1–4 interactions) are scaled by factors 1/2 (Lennard-Jones) and 1/1.2 (Coulomb).[22] The reason for such omitting 1-2 and 1-3 non-bonded interactions is understandable: during parametrization process the non-bonded interactions between atoms connected by bonds were effectively described by bonding (or bending) parameters and additional non-bonded interaction would lead to the wrong results.

Typically, for non-bonded interactions only the nearest pairs are considered. The cut-off distance is defined and all pair interactions behind this distance are omitted. This strategy works well only with fast-decaying interactions such as the dispersion one but for electrostatics different approach has to be applied. Particle-mesh Ewald (PME) summation[25,26] is the method of choice for long-range electrostatic interactions in most recent MD program packages. PME was used also in the presented thesis.

1.2.3 Dispersion

In the Amber force field the dispersion energy is covered by the Lennard-Jones potential, which can be written in two equivalent forms:

$$V_{ij}^{LJ} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{R_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}}\right)^{6}\right] \tag{1.9}$$

$$V_{ij}^{LJ} = \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \tag{1.10}$$

where ε_{ij} represents the depth of potential energy minimum and σ_{ij} is the value, where potential energy is equal to zero. Parameters A_{ij} and B_{ij} can be expressed as

$$A_{ij} = 4\varepsilon_{ij}\sigma_{ij}^{12} \tag{1.11}$$

$$B_{ij} = 4\varepsilon_{ij}\sigma_{ij}^{6}. \tag{1.12}$$

Probably from the pure laziness of the force field designers there are not pair parameters A_{ij} or ε_{ij} but atom parameters A_i and ε_i in the Amber parameters set. The pair parameters are calculated from the atom parameters applying various combination rules. Amber uses Loretz-Berthold combination rules[22,27] together with ε_i and σ_i atom parameters:

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \tag{1.13}$$

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)\,. \tag{1.14}$$

1.2.4 Force Field Modifications

The Lennard-Jones potential contains repulsion and dispersion (attraction) part. Despite the fact that the functional form of Lennard-Jones (Eqn. 1.9) does not clearly say which parameter should be scaled to cancel the attractive (dispersion) part, such scaling had been already used before.[28] In this thesis the other functional form (Eqn. 1.10) was used and the attraction was neglected by scaling term B_{ij}/R_{ij}^6 by factor 0.0 to mimic a loss of dispersion. As was mentioned, there are atom parameters ε and σ in the Amber force field. Consequently, a *library* of pair parameters A_{ij}/R_{ij}^{12} and B_{ij}/R_{ij}^6 was generated employing the correct combination rules.[22,27]

The systems of interest (solvated base pair or DNA) can be divided into two parts – solute (nucleic acid bases or DNA) and solvent (water). These to subsystems comprehend three types of non-bonded interactions, namely water–water (WW), water–solute (WS) and solute–solute (SS). The library of combined pair parameters allows us to choose which interaction we conserve and which we discard. A series of simulations was performed discarding different parts of non-bonded potential:

WW-WS-SS All interactions were conserved in unmodified fashion. This was a validation simulation where the Eqn. 1.10 functional for was compared with the simulation employing Eqn. 1.9 functional form (well tested default setup in the Gromacs program).

WW-WS Dispersion within the solute was neglect. However, for correct description of the torsions, 1-4 Lennard-Jones term was maintained. This might be

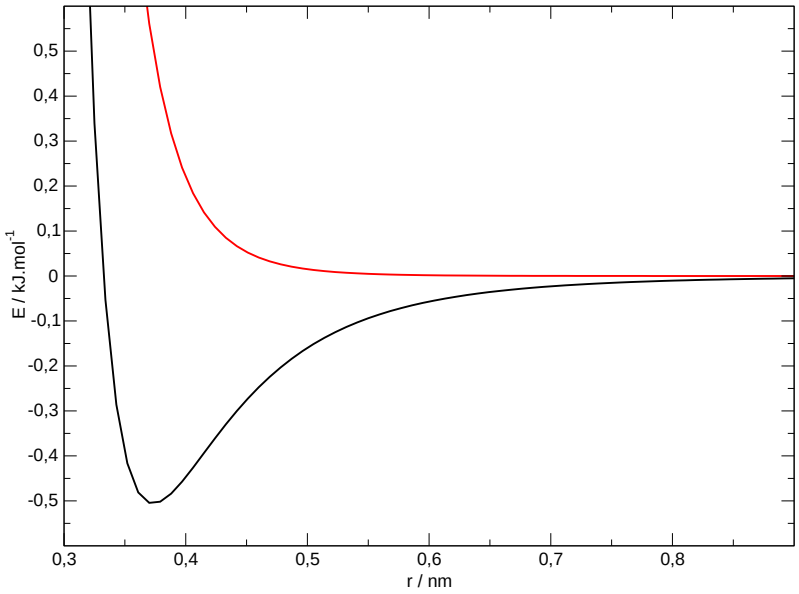


Figure 1.2: Lennard-Jones potential for atom types NB and CB defined in Amber force field. Black curve: $A_{ij}/R_{ij}^{12} - B_{ij}/R_{ij}^{6}$ and after modification red curve: A_{ij}/R_{ij}^{12}

important especially for DNA simulations, where the backbone torsions play significant role in dynamical behavior of the double-helix.[29,30]

WW-SS Dispersion between solute and water molecules was neglected.

WW Dispersion within solute as well as between solute and water was neglected.

The correspondence of the Lennard-Jones term B_{ij}/R_{ij}^{6} and the quantum chemical description of the dispersion interaction is fairly satisfying. The representation of the repulsion is, however unsatisfactory even in the unmodified case, when it grows with decreasing distance with r^{-12}, while it should grow rather exponentially. This form was chosen primarily for its computational simplicity and,fortunately, the parametrization improved its behavior towards the reality.

This repulsion term represents the effective atom size. While in the unmodified situation the atoms size results from the linear combination of r^{-6} and r^{-12} terms, in the modified situation only term r^{-12} plays the role. It is well known fact that the hydrophobicity as a property of a particle increases with the increase of the particle's surface area or volume.[31] In other words, hydrophobic effect (disruption of water

structure and dynamics) acts more strongly on the larger solutes than on the small ones. Thus this Lennard-Jones potential modification (neglecting of dispersion) brings a source of an inestimable error which is attributed to the increase of the effective size of the atoms. We can estimate the effective atom size from a radial distribution function of water molecules around the heavy atoms of the solute. In Fig. 1.3 there are radial distribution functions of heavy atoms of the Dickerson's DNA dodecamer in TIP3P water. The size of the atoms is proportional to the position of the first peak.

The simulation program, namely Gromacs *mdrun*, does not allow us to use different repulsion functional forms and if so, the parameters for DNA are not available. Partial solution might be a scaling of the A_{ij} parameters to obtain as much similar radial distribution function as possible. The position of the first peak has to be considered. The Fig. 1.3 shows how the radial distribution function changes with the scaling the A_{ij} parameter.

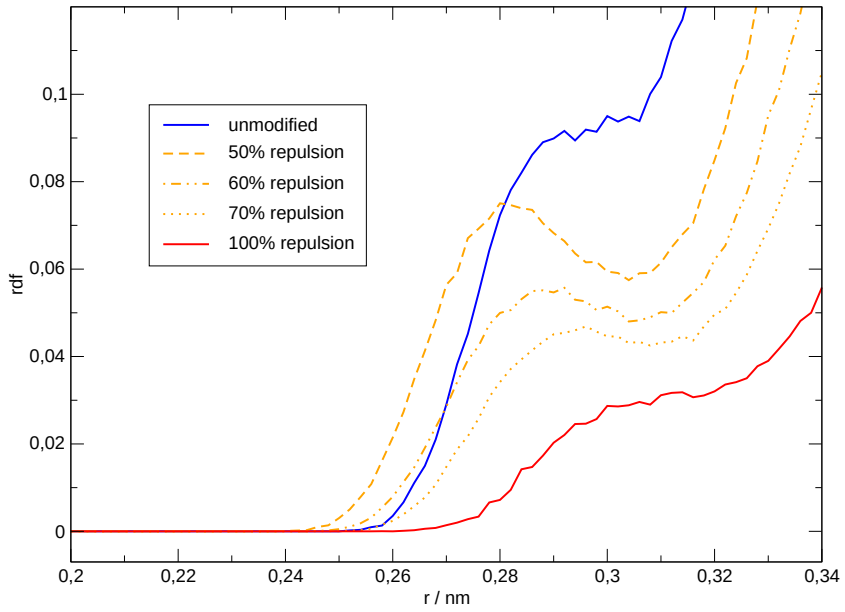


Figure 1.3: Radial distribution functions of unmodified WW-WS-SS (blue), modified WW (red) and modified WW with the scaled repulsion term to 50%, 60%, 70% (orange).

One has to keep in mind that the size varies and depends on the atom type. Scaling these parameters all together by one scaling factor to reproduce radial distribution function of water around a solute, makes some of them larger and some of them

smaller compared with the correct potential, but can assume the interaction between water and solute might to be effectively conserved, which should result in vanishing of the artificial increase of the hydrophobicity caused by neglection of dispersion.

1.3 System Representation

We studied two systems, 9-methyladenine and 1-methylthymine in water and DNA dodecamer in water.

1.3.1 Water

In the simulations explicit water model was adopted which means that each water molecule was represented separately. We used rigid three-center TIP3P water model introduced by Jorgensen,[32] where only oxygen atom has van der Waals parameters. This water model was designed to well reproduce experimental solvation energies and is less computational demanding compared to four-center models.[32] In TIP3P water model the atoms are kept in triangular shape by constraining algorithm SETTLE.[33]

1.3.2 Nucleic Acid Bases

The geometry of the bases was obtained by *ab initio* gradient optimization in Gaussian program[34] on the B3LYP/cc-pVTZ level. Around the optimized geometry the grid of electrostatic potential was calculated on the HF/6-31G* level and this was used for RESP fit of the atom charges[35]. For the fit the Antechamber program[36] was used. These charges are overestimated by about 15% since the empirical potential lacks polarization term, whose absence we want to compensate.[22]

Other parameters such as equilibrium bond lengths, force constants etc. were adopted from general Amber force field (June 2003).[37] This was done separately for 9-methyladenine and 1-methylthymine (Fig. 1.5). From the two bases we manually prepared stacked complex with respect to the gas phase *ab initio* optimized geometry.[38] The stacked complex was solvated in the rhombic dodecahedron box (Fig. 1.1) under the condition that the shortest distance between box and solute is 1.6 Å which resulted in box containing about 5000 TIP3P water molecules.

The solvated base pair was equilibrated by several steps:

1. minimization of the whole system (avoids possible close contacts)
2. 20 ps long water heating employing Berendsen thermostat[19] with fixed complex position under the constant volume and targeted temperature 300 K
3. 20 ps long heating of the whole system to 300 K employing Berendsen thermostat and barostat
4. 0.5 ns long equilibration run employing Nosé-Hoover thermostat and Parrinello-Rahmann barostat under the temperature 300 K and pressure 1 bar
5. 0.5 ns equilibration run with the umbrella potential (see Sec. 1.4.1) with small force constant 500 kJ.mol^{-1}.

After equilibration the final structure was taken as the starting one for umbrella sampling procedures. For the conversion from the Amber format to Gromacs format Ambgro program was used.[39]

1.3.3 Dickerson's DNA Dodecamer

The geometry of DNA dodecamer with sequence d(CGCGAATTCGCG) was taken from PDB database[40] (ID 1bna, 1.9 Å resolution). This high-quality X-ray structure was measured by Dickerson[41] in the late eighties and is one of the widely used DNA models.* Water molecules and ions were removed from the X-ray and hydrogen atoms were added. From the Amber library of residues all parameters were assign including also the non-bonded ones. We use Amber Parm99 force-field[22] which performs well also with the abnormal structures[42] and is able to describe conformational changes.[43] Hence the applicability of the force field is being find to be guaranteed for the presented purposes.

The double-helix was solvated in the rhombic dodecahedron box (Fig. 1.1) under the condition that the shortest distance between box and solute is 1.2 Å which resulted in box containing about 10000 TIP3P water molecules. From the bulk water it was

*it has 596 citations on the April 2009

chosen 22 molecules according to the electrostatic potential around the solute, and they were replaced by sodium cations to reach electroneutrality of the system.

The DNA surrounded by water was equilibrated in several steps:

1. minimization of water molecules applying position restraints with force constant 1000 kJ.mol^{-1} to the DNA to avoid possible close water contacts
2. minimization of the whole system
3. 20 ps long water heating to target temperature 300 K employing Berendsen thermostat with position restraints acting on DNA keeping the DNA under temperature 10 K
4. 20 ps long heating of the whole system to 300 K employing Berendsen thermostat and barostat
5. 1 ns long equilibration run employing Nosé-Hoover thermostat and Parrinello-Rahmann barostat under the temperature 300 K and pressure 1 bar

The final structure from the equilibration procedure was taken as the starting structure for other DNA simulations.

1.4 Free Energy Calculations

The free energy calculations represent very challenging part of the computational chemistry. Classical thermodynamics says that the spontaneous changes in the nature are connected with the negative change of the free energy. Statistical thermodynamics gives us some theoretical insight into microscopical correspondence of the free energy and fundament of statistical mechanics – partition function. The relationship between both can be expressed as

$$F = -k_B T \ln Q \tag{1.15}$$

where F stands for Helmholtz free energy, k_B is Boltzmann constant, T is temperature and Q is canonical partition function. Unfortunately, partition function is (similar to wave function) unreachable in exact form and we need to approximate it.

1.4.1 Umbrella Sampling

There are several methodologies how to sample configurational (or phase) space to obtain the estimate of the free energy change.[44] We employed so called *umbrella sampling* method firstly introduced by Torrie et al[45] in the Monte Carlo simulations. The main problem in the molecular dynamics sampling of configurational space that the system is trapped in some deep potential energy minimum and the barriers are overcome with very low probability, is in umbrella sampling overcome by biasing the simulation by an additional potential added into potential energy function (*c.f.* Fig 1.4).

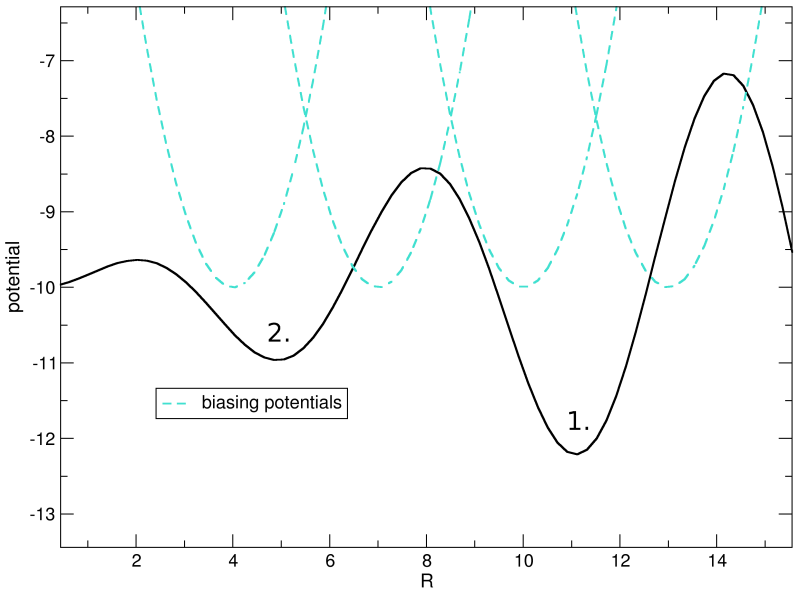


Figure 1.4: General umbrella sampling scheme. Without bias (black curve) the system is trapped in region 1. or 2. due to the potential energy minima and the barrier between regions 1. and 2. is in the unbiased simulation sampled poorly. Sampling can be improved by biasing potential. Four umbrella sampling windows are shown, each one contains a harmonic potential centered on different value of R.

It has to be mentioned that there is no physical correspondence of the biased system in real world but from this unphysical situation we try to conclude, after some mathematical manipulation, clearly physical results.

The free energy along some coordinate $\vec{R}$ is connected with a probability distribution

along the coordinate $P(\vec{R})$ by the equation

$$F(\vec{R}) = -k_B T \ln P(\vec{R}). \tag{1.16}$$

If one introduces an additional potential V_{bias} into the simulation the biased probability distribution $P'(\vec{R})$ is obtained and the free energy can be evaluated as follows

$$F(\vec{R}) = -k_B T \ln P'(\vec{R}) - V_{bias} + C \tag{1.17}$$

where C is an arbitrary constant, which is unknown, but since we are interested in free energy changes it is irrelevant. The situation is changed, when the sampling along the coordinate is not improved enough. Then one solution is to divide the reaction coordinate into several windows and in each window biasing potentials are introduced (Fig. 1.4) differing usually in the equilibration value of the coordinate. The free energy profile is then obtained by overlapping the partial probabilities. In such case, the constants C are different in each window, cause an undetermined offset of neighbouring windows, can not be simply ignored and advanced technique has to be employed to evaluate the unbiased free energy profile.

1.4.2 Weighted Histogram Analysis Method (WHAM)

Weighted histogram analysis method developed by Kumar[46] provides unbiased probability distribution from biased probability distributions (histograms) of each simulation window by minimizing the constants' C differences. The *WHAM* equations

$$P(R) = \frac{\sum_{i=1}^{N_{win}} n_i(R)}{\sum_{i=1}^{N_{win}} n_i e^{\frac{C_i - V_{bias,i}(R)}{k_B T}}} \tag{1.18}$$

$$C_i = -k_B T \ln \sum_{R_{bins}} P(R) e^{\frac{-V_{bias,i}}{k_B T}} \tag{1.19}$$

where $P(R)$ represents the best estimate of unbiased probability distribution, N_{win} is number of simulation windows, n_i is number of counts in the bin associated with R,

C_i is the constant in the *i-th* window and $V_{bias,i}$ represents biasing potential of *i-th* window, has to be solved iteratively, due to the cross-dependence of $P(R)$ and C_i.

1.4.3 Potential of Mean Force

The free energy change can also be evaluated on the bases of different idea, so called *potential of mean force* concept, firstly mentioned by Kirkwood.[47] It exploits the fact that the change of arbitrary function $q(x)$ can be expressed as follows

$$\Delta q(x) = \int_a^b \frac{\partial q}{\partial x} dx. \tag{1.20}$$

It can be further shown that

$$\frac{\partial G}{\partial x} = \left\langle \frac{\partial V(x)}{\partial x} \right\rangle = \langle -f(x) \rangle \tag{1.21}$$

where $V(x)$ is the potential energy, $f(x)$ is general force and brackets $\langle\rangle$ denote averaging over the MD frames. From the total potential energy function V it can be extracted the umbrella potential $V_{bias}(R)$, which in case of harmonic form gives the force as

$$f(R) = \frac{\partial V_{bias}}{\partial R} = -k_f \left(R - R_0\right) \tag{1.22}$$

where k_f is a force constant and R_0 an equilibrium distance which the both are the parameters describing the biasing harmonic potential. Simple averaging over the frames of each window and consecutive integration gives us potential of mean force which describes changes in free energy. Let us add that compared with *WHAM* evaluation which uses data from all windows, this naive evaluation considers only data from one window to calculate the force.

1.4.4 Entropic Correction

When the reaction coordinate is a distance between two particles, the free energy in larger distances would decrease. It is caused by entropic contribution. We can

imagine one particle fixed during elongation of the distance and second particle moving on the sphere surface around the first one. Since the surface increases with the square of the distance, the increasing entropy connected with the larger number of sub-states in the configurational space causes the decrease of free energy. The potential of mean force of two non-interacting particles can be expressed as

$$PMF = -(n_c - 1)k_B T \ln R \tag{1.23}$$

where n_c represents the dimensionality of the biasing potential.[14] Since the free energy of the association/dissociation of the non-interacting particles should be zero, this term covers exactly the entropic part, which has to be subtracted also from the other resulting *pmf* curves.

1.4.5 Simulation Setup

As a reaction coordinate describing the association/dissociation of the nucleic acid base-pair, the distance between two heavy atoms from the centers of the molecules (highlighted in Fig. 1.5) was chosen.

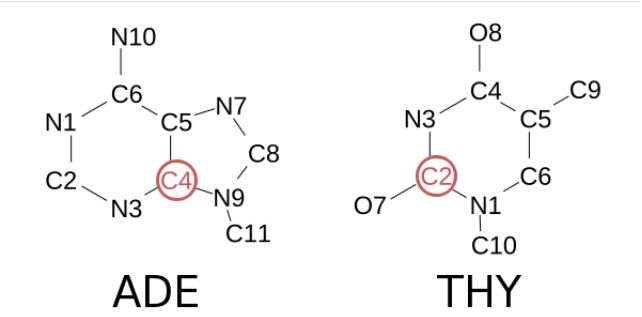


Figure 1.5: Atoms numbering of 9-methyladenine and 1-methylthymine. For clarity hydrogen atoms were omitted.

Also other pairs of atoms were checked, but as is discussed later, no significant dependencies of the free energy on the choice of the pair of atoms were observed. The distance was varied in the range between 0.3 nm and 2.0 nm with the step 0.05 nm. Consequently 34 simulation windows were generated differing in the biasing potential. Force constant was adjusted to 1000 kJ.mol^{-1} which is rather *soft*

harmonic potential suitable more for *WHAM* evaluation than for evaluation based on the forces calculation (Sec. 1.4.3).

The starting structure of each simulation window was taken from the previous window but velocities were generated randomly according to the Maxwell distribution within 300 K. Equilibration part 500 ps long was followed by 3.5 ns production run. The base pair structure was saved every 0.1 ps providing us 17500 data points per each window.

1.5 Solvent Accessible Surface Area

The concept of solvent accessible surface area was firstly proposed by Lee[48] in 1971. It is a measure of the size of a molecule (*e.g.* protein or hydrocarbon).

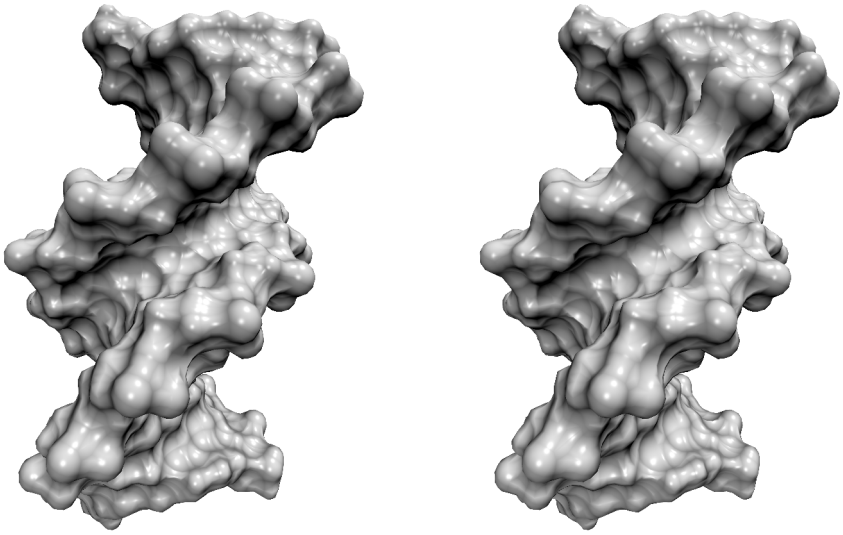


Figure 1.6: Solvent accessible surface areas of Dickerson's dodecamer calculated with probe radius 0.15 nm (left) and 0.17 nm (right). Apparently, there are only subtle visible distinctions.

The idea is to approximate the water molecule by a *probe* sphere with some van der Waals radius (as is described *e.g.* in TIP3P water model) and roll the probe sphere around the molecule obtaining a molecular surface *e.g.* as shown in Fig. 1.6 . Usually, the probe radius is set to the value of 0.15 nm (σ parameter of oxygen atom in TIP3P water model is 0.158 nm) but for our purposes we used 0.17 nm (results are explained later). Due to the larger probe radius the absolute values we calculated do not exactly

represent solvent accessible surface areas for water solvent but they serves merely as some characteristic of the DNA.

1.6 Helical Parameters

As a validation descriptors for the DNA simulations we used the helical parameters introduced by Dickerson *et al*[49] and implemented in X3DNA program package.[50] For our purposes only the subset which consists of the *base-pair step* parameters and the *complementary base-pair* parameters were considered. The trajectory of the DNA with unmodified potential using both Lennard-Jones functional forms (Eqn. 1.9 and Eqn. 1.10) was analyzed every 10 ps and the average of each parameter over all base-pairs was taken as the characteristic.

The terminal nucleotides are significantly more flexible than the rest of the double-helix. Thus they were not involved in the trajectory analysis.

Chapter 2

Nucleic Acid Base-Pair

2.1 Introduction

To get some insight into the role of dispersion in DNA stabilization we decided to perform the free energy calculations involving those DNA building blocs, which we believe are strongly influenced by dispersion interaction. Those building blocks – the nucleic acid bases – are quite common systems for modelling interactions participating in stabilization of biological structures. There have been done huge amount of calculations concerning DNA bases.[38,51–53] Among them one can find several papers estimating the free energy changes during association/dissociation or the solvation free energies.[52,54–56] Unfortunately, the papers provide notably different numbers depending on which methodology was used which reflects the fact that to estimate the free energy accurately, moreover when the changes are in a scale of single kcal, seem to be a very difficult task even nowadays.

The methodology we used provides us a nice possibility how to control the pair nonbonded interactions within the system and as we hope it enables passing the quantitative results to the DNA case.

2.2 Free Energy Profiles

At the beginning of this section some benchmark calculations, by which we justify the chosen methodology for evaluation of the free energy profiles, are introduces. In

the second part the set of simulations based on modified Lennard-Jones potentials is presented and discussed. The profiles were shifted according to the zero reference level of the separated bases.

2.2.1 Testing Calculations

On Fig. 2.1 (left hand side plot) the validation of the pair parameters library is shown. There are two curves obtained with unmodified Lennard-Jones potential with Eqn. 1.9 and Eqn. 1.10 functional form, respectively. Any deviations can only be caused by numerical uncertainties and/or partially unconverged results.

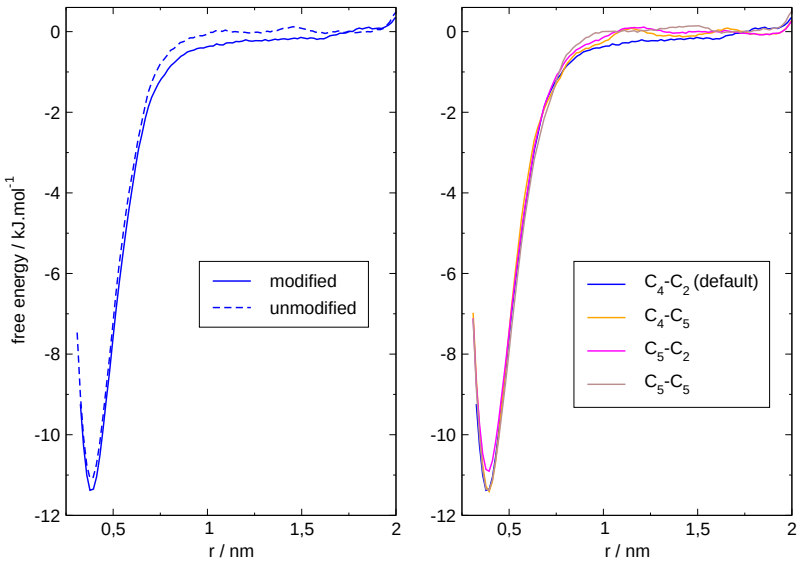


Figure 2.1: LEFT: Comparison of Eqn. 1.9 and Eqn. 1.10 functional forms of Lennard-Jones potential *i.e.* validation of the pair parameters library. RIGHT: Comparison of different atom pairs describing the reaction coordinate of the association/dissociation.

The dependence of the free energy profile on the choice of atoms pair is shown on Fig. 2.1 (right hand side plot). Several different pairs for distance coordinate definition were tested. These were C_4-C_5, C_5-C_2 and C_5-C_5 meaning the first atom from methyladenine and the second atom from methylthymine (*c.f.* Fig 1.5). The observable differences are within the statistical error of the simulations, thus the curves

can be considered as identical and the free energy profile independent on the choice of the reaction coordinate from the range of the tested ones.

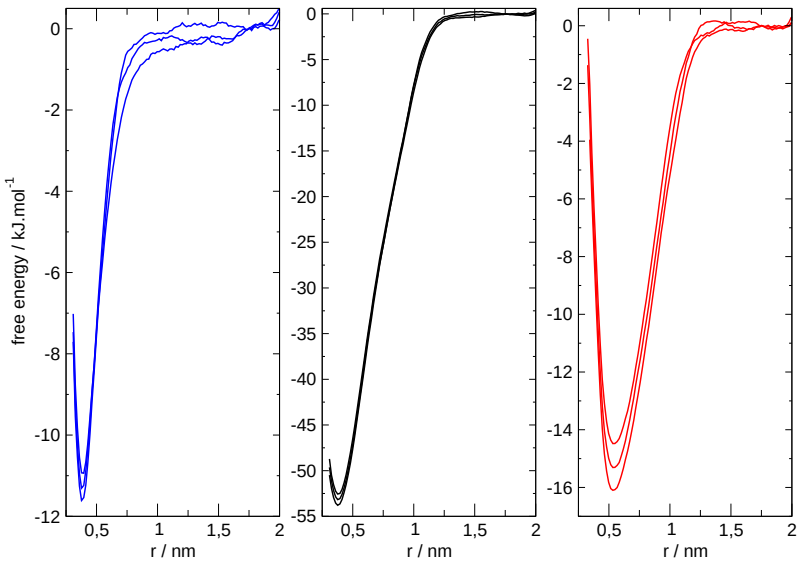


Figure 2.2: Convergence performance. Plotted profiles are from the first, second and third nanosecond of the WW-WS-SS (blue), WW-SS (black) and WW (red) force-field modifications.

The free energy convergence was checked by partial evaluation of the umbrella sampling windows. The free energy profiles obtained from *WHAM* evaluation of 1 ns long segments of each window are plotted in Fig. 2.2 for three consequential segments. The profiles are satisfyingly good converged with the largest variance of 1.6 kJ.mol^{-1} for WW (red) case. Critical appears to be the region with the distance around 0.8 nm, where the structures are getting to be floppy and the simulation has to be long enough to sufficiently sample those situations. For umbrella sampling method evaluated by *WHAM* procedure the overlap of the neighbouring windows is important. Fig. 2.3 shows that histograms (unbiased probability distributions) for all umbrella windows for WW-WS-SS case overlap sufficiently.

For comparison, the curves obtained by potential of mean force methodology (see Sec. 1.4.3) are shown in Fig. 2.4. We present WW-WS-SS (blue) WW-WS (black) and WW (red) curves. Clearly, all the curves are identical considering their shape. As was mentioned in the Methods section, the *pmf* evaluation uses data from each window separately due to the force averaging over the one particular window's frames,

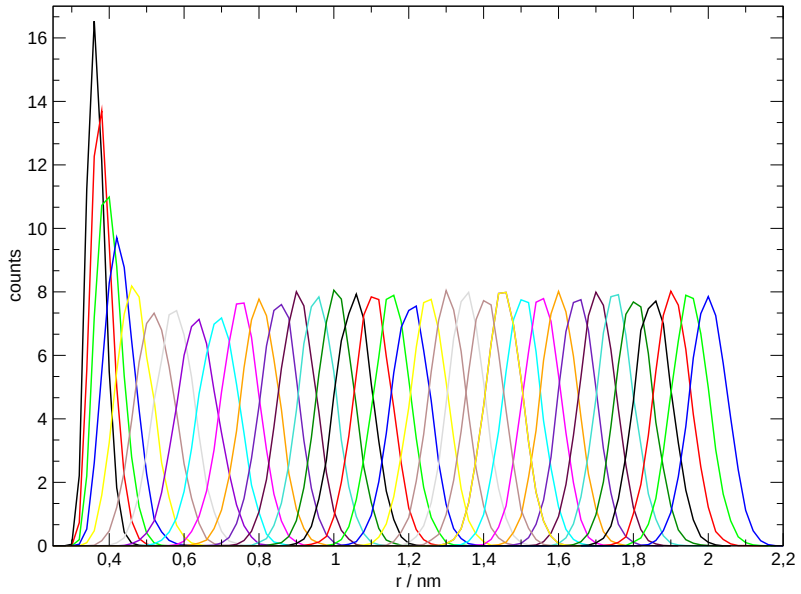


Figure 2.3: Histograms for the unmodified simulation (WW-WS-SS).

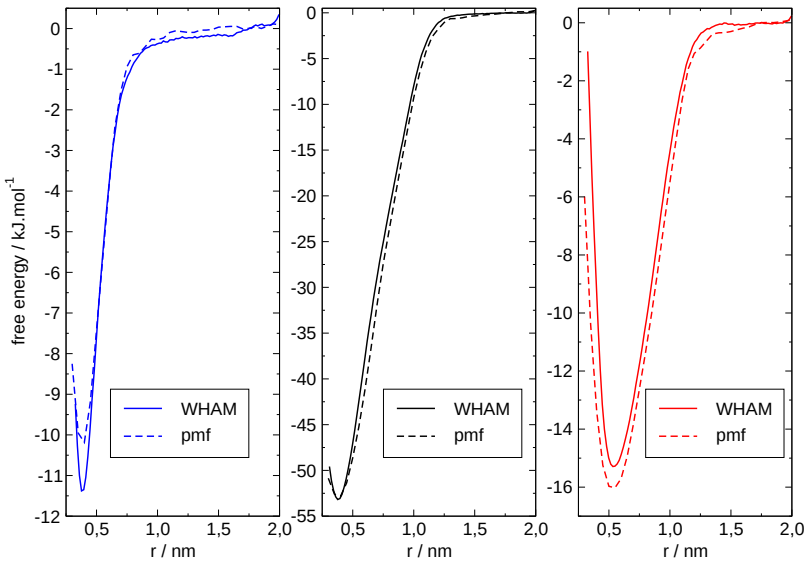


Figure 2.4: Comparison of *WHAM* and *pmf* evaluations for three different simulations – WW-WS-SS (blue curves), WW-SS (black curves) and WW (red curves)

thus should be considered as the less precise than the *WHAM* evaluation, which uses for generation of the free energy profile all available data across the whole reaction coordinate.

2.2.2 Results

Fig. 2.5 shows the free energy profiles for all four situations: WW-WS-SS (blue), WW-WS (green), WW-SS (black) and WW (red). These curves were obtained by umbrella sampling of the distance coordinate between mA-C_4 and mT-C_2 atoms later evaluated with *WHAM* procedure (see Sec. 1.4.2).

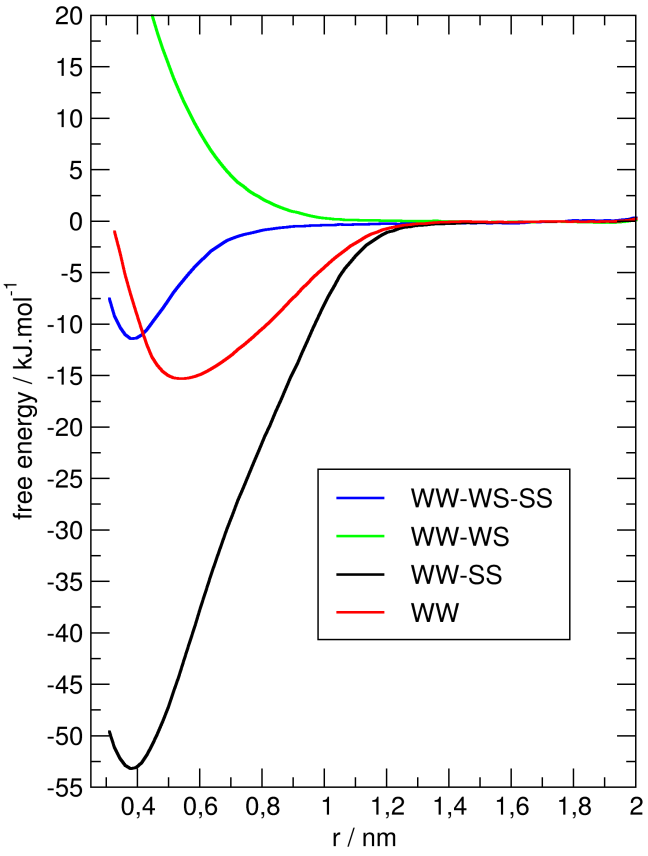


Figure 2.5: Free energy profiles for unmodified and three modified Lennard-Jones potentials.

WW-WS-SS The blue curve in Fig. 2.5 represents unmodified case, which means that all Lennard-Jones interactions were conserve across the whole system. The minimum located at the C-C distance slightly under 0.4 nm indicates that the complex is in the stacked structure. No evidence for hydrogen-bonded structure whose minimum would be located around 0.8 nm is present. All the hydrogen-bonded structures are short-lived which can not cause noticeable decrease of the actual free energy. This is completely different situation than in the gas phase, where hydrogen-bonded structures are preferred.[38] The missing hydrogen-bonded complex on the free energy profile is in agreement with previous studies[56–60] which conclude that the stacked structures of different aromatic molecules such as purines or pyrimidines are preferred in water environment and hydrogen-bonded complexes are not thermodynamically stable in water.

This can be explained by the much worse hydration of the hydrogen-bonded structure compared with the stacked one. As discussed by Chandler[61] and later in this thesis, hydrophobic effect in the classical sense of disruption of water structure and dynamics leads the bases into more compact stacked structure rather than into the hydrogen-bonded one despite the large favourable enthalpic contribution of the H-bonds between bases.

The value of –11 kJ.mol^{-1} of association free energy of the studied bases is in the widely accepted range of values of substituted purines and pyrimidines, which is from already mentioned reasons quite wide. Theoretical and experimental values[52,56,62–64] vary between –3 to –23 kJ.mol.$^{-1}$ However, for the purposes of this study the relative differences between simulations with different conserved amount of Lennard-Jones interaction are important rather than the absolute values.

WW-WS The green curve in Fig. 2.5 represents the free energy profile of association/dissociation, where dispersion interaction between solute (nucleic acid bases) and water was neglected. It has to be mentioned that the Lennard-Jones interaction decays very quickly (with r^{-6}) and therefore the bases would not interact in long distances. Hence when we discard one type of interaction, which expresses only in short distances (that means only on one side of the reaction *base + base* $\rightarrow$ *complex*) the equilibrium has to be consequently shifted to large distances (that means the *base + base* side of the reaction). We should not be surprised, that the minimum on the green curve is not present. From the comparison with blue curve we can

conclude that the dispersion interaction between bases supports complex formation.

WW-SS Complementary to the previous case is the black curve on Fig. 2.5 where dispersion interaction between solute and water was cancelled. Contrary to the previous case the dispersion energy between bases and water goes against the complex formation which can be understood from the deeper minimum of the black curve compared with the blue one. The minimum of the free energy is –53 kJ.mol.$^{-1}$

WW The last case is the situation where dispersion interaction of the bases is completely discarded which means that there is no dispersion interaction between bases nor between bases and water. The free energy profile is represented by the red curve in Fig. 2.5. The depth of the minimum is –15 kJ.mol.$^{-1}$

If we understand well the green and black free energy curves than the red curve lacks two opposite effects – complex stabilising base–base dispersion interaction and complex destabilising base–water dispersion interaction. The red curve exhibits significant minimum, subtly shifted to larger distance and deeper compared with the blue curve. There are several possible reasons which contribute to such free energy profile shape.

By comparing the blue and black curves with the red one we can considered the water–solute dispersion energy contribution and the solute–solute dispersion energy contribution being unequal with the water–solute one being larger in the absolute values.

Also the increase of solute entropy might play a role. At first glance seen, the significant distinction of the red and blue curves is that the former one is much broader. This reflects the larger variability of the complex structures. This larger variability of the complex seems to be an important characteristic and the higher entropy of the nucleic acid bases can result. The unmodified case amounts to lower entropy, since the dispersion interaction between bases restricts the configurational space accessible for the system. The trajectory shows the complex structure without dispersion to be very floppy and plenty of stacked-like and T-shape-like structures with larger intersystem distance occurs.

Probably the last possible contribution to the deeper minimum on the red curve might be the artificial increase of the hydrophobicity of the nucleic acid bases (see Sec. 1.2.4). Neglecting the dispersion term in Lennard-Jones potential we enlarge the effective

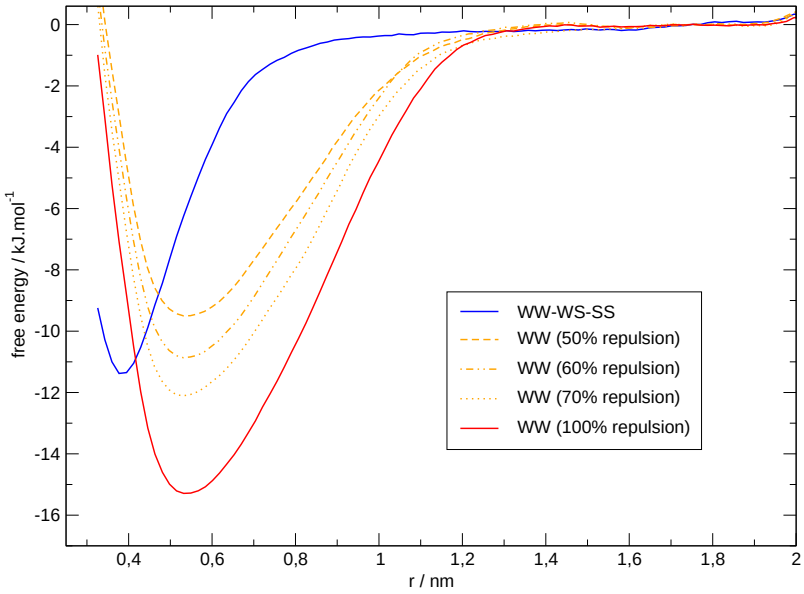


Figure 2.6: Free energy profiles of WW-WS-SS (blue), WW (red) and WW with the scaled repulsion term (orange curves).

atom size (see Fig. 1.3). This unwanted effect is caused by the Lennard-Jones formulation, which is currently the most common one. Better description of non-bonded interactions is necessary, but nowadays unavailable together with the flexibility needed for manipulations similar to that, in this thesis.

To vanish the artificial hydrophobicity contribution we scaled A_{ij} repulsion parameter in Lennard-Jones potential by factors 0.5, 0.6 and 0.7. Figure 2.6 shows the free energy profiles for unmodified WW-WS-SS (blue) case, red WW case and orange WW cases with scaled repulsion. It's clear, that the solute size strongly affects the depth of the free energy minimum. From the radial distribution functions we concluded that scaling repulsion by factor 0.6 gives similar solute size as with the unmodified potential.

The red curve in Fig. 2.6 with repulsion scaled to 60% has much the same depth as the unmodified case, from which one could understand that the water–solute and solute–solute contributions are comparable in absolute values. This finding is important and is later used in the second part concerning the simulations of DNA.

Chapter 3

Dickerson's Dodecamer

3.1 Introduction

In this chapter the results of the simulations of the Dickerson's DNA dodecamer (see Sec. 1.3.3) are presented. We performed a series of simulations with the same Lennard-Jones potential modifications as in the nucleic acid bases case and the DNA simulations are differentiated with the same colours as the bases are. The WW-WS (black) modification was omitted from the DNA calculations since it does not promise any development in the problem understanding.

3.2 Results

WW-WS-SS The blue row in Fig. 3.1 illustrates several snapshots from the DNA simulation, where no modifications of Lennard-Jones potential were accomplished.

During the simulation no disruptive changes in the structure appeared. Verification of the use of different functional form of Lennard-Jones potential (Eqn. 1.10) was done by comparison of the helical parameters (see Sec. 1.6) with those obtained from the simulation based on Eqn. 1.9 and with those from crystal structures.[65] The helical parameters are shown in Tabs. 3.1 and 3.2. The reasonable agreement between the results obtained with different functional forms is evident but there are some distinctions in comparison with the experimental values. Since the calculated values represent averages over the complete sequence and the parameters are sequence

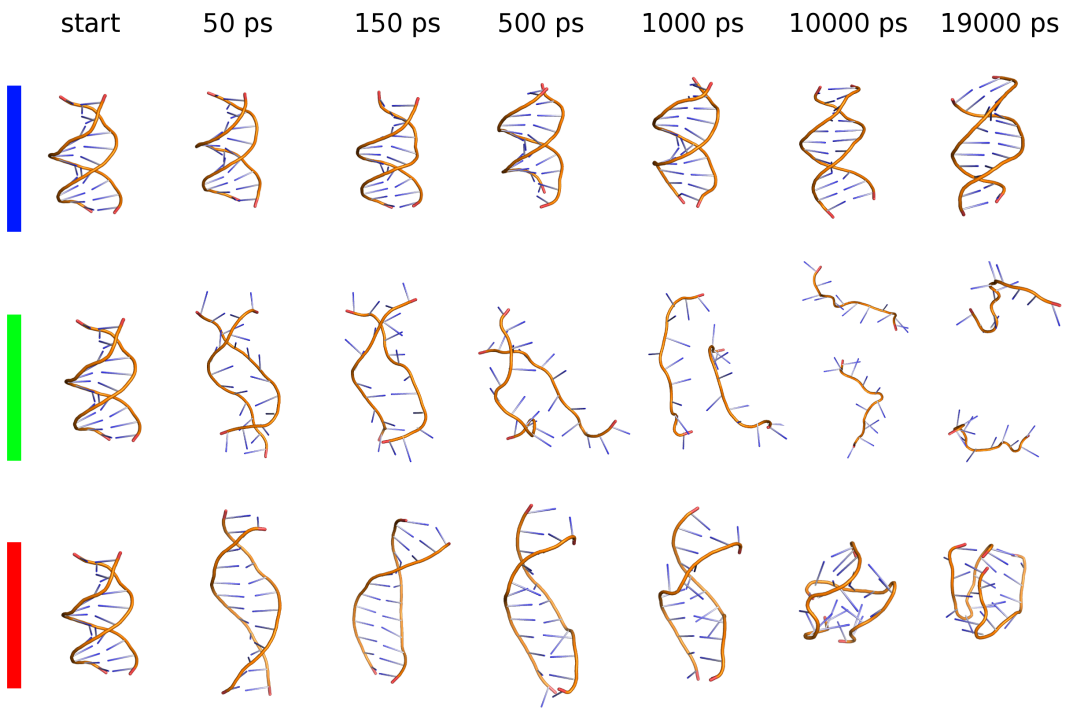


Figure 3.1: Snapshots of the DNA dynamics. The colours correspond with the free energy profiles of nucleic acid bases association/dissociation.

dependent, the discrepancies might relate with the fact that Dickerson's dodecamer's sequence does not cover all possible pairs. In addition, Amber force-field is known to underestimate helical twist.[66,67] The alternative functional form together with the *library* of pair A_{ij} and B_{ij} parameters can eventually be considered as reliable.

WW-WS The green row in Fig. 3.1 illustrates dissociation of two DNA strand if the dispersion energy within DNA was neglected whereas the dispersion interaction between DNA and water was conserved. The hydrogen bonds between the DNA strands were cleaved immediately which resulted in the DNA separation. Due to the attractive Lennard-Jones interactions between strands and water there was no will of the strands to fold and their complete expansion into bulk water was observed during the first nanosecond. The solvent accessible surface area is shown in Fig. 3.2 (discussion bellow).

Accordingly to the explanation of the green free energy profile (Fig. 2.5), the lack of dispersion energy within the DNA duplex moves the equilibrium to the single-

	alt.*	val.†	exp.‡
buckle	0.32	-0.10	0.50
propeller	-12.00	-12.18	-11.40
opening	-0.14	-0.27	0.60
shear	0.01	0.00	0.00
stretch	-0.04	-0.04	-0.15
stagger	0.06	0.08	0.09

Table 3.1: Complementary base-pair helical parameters. * alternative functional form for LJ potential (Eqn. 1.10), † default functional form for LJ potential (Eqn. 1.9), ‡ crystal structures from Ref.[65]

	alt.*	val.†	exp.‡
tilt	-0.12	-0.04	-0.1
roll	2.68	2.49	0.6
twist	32.54	32.73	36.0
shift	-0.01	-0.02	-0.02
slide	-0.51	-0.48	0.23
rise	3.35	3.34	3.32

Table 3.2: Step base-pair helical parameters. * alternative functional form for LJ potential (Eqn. 1.10), † default functional form for LJ potential (Eqn. 1.9), ‡ crystal structures from Ref.[65]

strands-side because the DNA–water interaction becomes more favourable. The extremely high values of the solvent accessible surface area (Fig. 3.2) for green WW-WS case are connected with the expansion of the separated strands into the bulk water.

WW This, the most interesting situation if the dispersion interaction within the DNA duplex as well as between DNA and water is neglected, seems to be also the most complicated one. The simulation, where only dispersion energy among water was conserved, provided a couple of bizarre structures.

During the first 100 ps (the red row in Fig. 3.1) the DNA elongated in the direction of the helical axe and passed into a ladder-like structure. In this structure all hydrogen bonds were maintained, but their larger fluctuations were present. Solvent accessible surface area raised significantly. The fluctuations of the nucleic acid bases led to the cleavage of the hydrogen bonds between them which was followed by folding process of the harmed DNA structure. This folding was finished after 1–2 ns and

once the DNA was folded it kept compact structure with only some restricted motions. Large structure reorganisations were not during 20 ns long simulation observed, particularly due to the high energy barriers within the fold or due to the longer time scale required for such reorganisation.

The folded structure is characterized by smaller amount of hydrogen bonds compared with the double-helix form but the intrastrand hydrogen bonds occur too. There are also cavities inside the folded structure which are small enough to contain some water molecule. This fact makes the algorithm for solvent accessible surface area calculations useless because we can see from the trajectory that actually no water molecules are presented inside these cavities. Therefore we calculated solvent accessible surface area with a bigger probe molecule (0.17 nm), which does not fit into the cavities and the algorithm is able to give us for all Lennard-Jones modifications comparable results, which are eventually more illustrative.

The resulting solvent accessible surface area is lower than in the ladder-like structure but, however, higher compared with the double-helix (see Fig. 3.2).

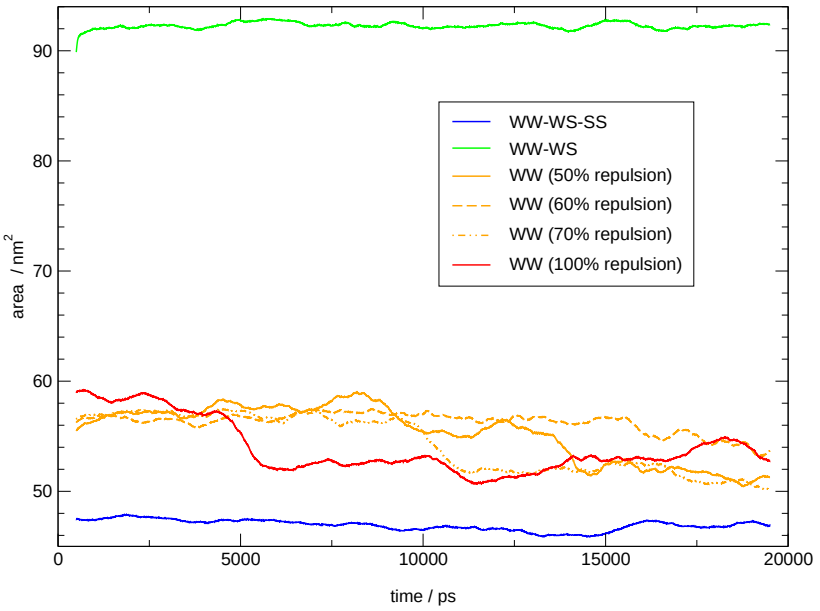


Figure 3.2: Dependence of the solvent accessible surface area on the simulation time. See Sec. 1.5 for details.

3.3 Hydrophobicity of the Folded DNA

As already mentioned, the potential energy function without dispersion term suffers from unwanted changes also in the repulsion description. It is a question what the leading force of DNA folding with the red Lenard-Jones potential modification is.

Since the folding runs relatively quickly the equilibrium is maintained in 20 ns long simulation. This would suggest the thermodynamical stability of the fold contrary the double-helix.

A set of additional simulations was performed to uncover the effect of the artificial hydrophobicity increase. We scaled A_{ij} repulsion term in Lennard-Jones potential by factors 0.5, 0.6 and 0.7 and evaluated the radial distribution functions of water around the heavy atoms of the DNA. This modification led to the reduction of the effective van der Waals radius (atom *size*) which is obvious from the Fig. 1.3 and factor 0.6 provided the radial distribution function nearly identical with the unmodified one with respect to the position of the first peak. It means that hydrophobicity of the original DNA (with unmodified potential) and hydrophobicity of the DNA obtained after modification (no dispersion and 60% of repulsion) are comparable.

The resulted structures after 20 ns long WW simulations with scaled repulsion term are shown in Fig. 3.3. All three DNA with various magnitudes of the repulsion folded into the structures easily distinguishable from double-helix. Moreover, the folding process ran through the ladder-like structure which were more or less flexible. The flexibility of the nucleic acid bases seems to be crucial in the folding process. The structure with 60% of repulsion tended to remain in the ladder-like structure supported by hydrogen bonds between bases for relatively long time. Once the hydrogen bonds were cleaved, the folding process could start. Thus the ladder-like structure can be considered as a kinetically metastable.

Considering the orange curves in Fig. 3.2 the tendency of the hydrophobic effect to pack the unwound DNA is undoubted. Let us remark that this is the pure hydrophobic effect in the sense of minimization of the solute size which leads to decrease of disruption of water structure and dynamics.

The larger effective atoms size in the WW (red) case can also explain the presence of cavities in the folded structure large enough for a water molecule but without

Figure 3.3: Final folded structures of DNA after 20 ns long simulation with various scaling of the repulsion term. From left hand side 50% repulsion, 60% repulsion and 70% repulsion. Dispersion interaction discarded in total (red WW case). The middle structure after 20 ns preserved a domain which is still ladder-like/hydrogen-bonded.

any evidence of that in the simulation. Other explanation of the cavities takes into account the restricted flexibility of the DNA sugar-phosphate backbone.

Chapter 4

Conclusions

4.1 Nucleic Acid Bases

1. In the water environment 9-methyladenine and 1-methylthymine create stable stacked complex The free energy of association amounts of -11 ± 0.6 kJ.mol^{-1}.

2. Modifying the Lennard-Jones potential in such way that the dispersion term between nucleic acid bases is discarded, the minimum on the free energy curve disappears. This contribution of dispersion energy is thus favourable for complex formation.

3. Discarding the dispersion energy between water and solute (nucleic acid bases) the minimum on the free energy curve becomes significantly deeper. We can conclude that the dispersion between water and solute *destabilizes* the complex stacked structure.

4. When both previously mentioned contributions absent together, in other words there is no dispersion interaction between nucleic acid bases nor between bases and water, the minimum on the free energy curve becomes deeper by about 4 kJ.mol^{-1} but after correction for atom size, the depth is comparable with the real (unmodified) case. These leads us to the conclusion that the base–base dispersion interaction, which is stabilising with respect to the free energy, is in absolute value comparable with destabilising bases–water dispersion interaction. Furthermore, the force which is responsible for the minimum on the free

energy curve seems to be the hydrophobicity in the classical meaning. The separated bases disrupt the water structure more significantly than the complex stacked structure. This effect is, however, not able to keep as compact structure as in the unmodified simulation is kept by the combination of several effects.

5. We thus conclude that in the real world the hydrophobicity appears to be the dominant factor stabilising stacked structures of aromatic solutes in water.

4.2 Dickerson's DNA Dodecamer

1. Discarding the dispersion within DNA atoms the double-helix dissociates into single-strands which are expanded into bulk water. The reason is the more favourable interaction between water and DNA, which is present in total, compared with interaction within DNA, which lacks the dispersion part.

2. When the dispersion interaction of DNA is neglected at all, which means there is no dispersion interaction between DNA and water nor within DNA itself, the DNA goes through ladder-like structure into folded structure being pushed forward by hydrophobic interaction.

   After correction for atoms size the folding became slower due to the larger stability of the ladder-like structure. This kinetically stable structure turns into folded one after initialization caused by cleavage of the hydrogen bonds between nucleic acid bases.

3. The consequence with reality is analogous to nucleic acid base one: the *stabilising* effect of hydrophobicity is emphasized over the dispersion interaction contribution, which is rather responsible for the maintaining (or recognition) of the double-helical structure. The packing of the two strands is caused by hydrophobic effect which is, however, not able to distinguish between the biologically relevant (and without questions real) double-helix structure and the hypothetical folded one, whose biological activity would be questionable. The dispersion interaction together with hydrogen bonds play somewhat an essential role as the *structure-shaping* factors.

References

[1] J. Watson and F. Crick, "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.", *Nature* **171**(4356), pp. 737 (1953).

[2] J. Watson and F. Crick, "Genetical implications of the structure of deoxyribonucleic acid.", *Nature* **171**(4356), pp. 964 (1953).

[3] P. Hobza and J. Šponer, "Structure, energetics, and dynamics of the nucleic acid base pairs: Nonempirical ab initio calculations", *Chemical Reviews* **99**(11), pp. 3247–3276 (1999).

[4] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, "Verification of the crooks fluctuation theorem and recovery of RNA folding free energies", *Nature* **437**(7056), pp. 231–234 (2005).

[5] J. Černý and P. Hobza, "Non-covalent interactions in biomacromolecules", *Physical Chemistry Chemical Physics* **9**(39), pp. 5291–5303 (2007).

[6] R. Malham, S. Johnstone, R. J. Bingham, E. Barratt, S. E. V. Phillips, C. A. Laughton, and S. W. Homans, "Strong solute – solute dispersive interactions in a protein – ligand complex", *Journal of American Chemical Society* **127**(48), pp. 17061–17067 (2005).

[7] http://www.ch.cam.ac.uk/c2k/cj/comp.html.

[8] The Royal Swedish Academy of Sciences (1998), http://nobelprize.org/.

[9] "Foldinghome distributed computing", http://folding.stanford.edu/.

[10] J. Aqvist and V. Luzhkov, "Ion permeation mechanism of the potassium channel", *Nature* **404**(6780), pp. 881–884 (2000).

[11] V. Buch, A. Milet, R. Vácha, P. Jungwirth, and J. P. Devlin, "Water surface is acidic", *Proceedings of the National Academy of Sciences of the United States of America* **104**(18), pp. 7342–7347 (2007).

[12] L. Verlet, "computer experiments on classical fluids .i. thermodynamical properties of lennard-jones molecules", *Physical Review* **159**(1), pp. 98–& (1967).

[13] R. W. Hockney, S. P. Goel, and J. Eastwood, "Quiet highresolution computer models of a plasma.", *Journal of Computational Physics* **14**, pp. 148–158 (1974).

[14] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, and H. J. C. Berendsen, *Gromacs User Manual Version 4.0*, (2005).

[15] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations", *Journal of Computational Chemistry* **18**(12), pp. 1463–1472 (1997).

[16] B. Hess, "P-LINCS: A parallel linear constraint solver for molecular simulation", *Journal of Chemical Theory and Computation* **4**(1), pp. 116–122 (2008).

[17] S. Nosé, "A molecular-dynamics method for simulations in the canonical ensemble", *Molecular Physics* **52**(2), pp. 255–268 (1984).

[18] W. G. Hoover, "Canonical dynamics - equilibrium phase-space distributions", *Physical Review A* **31**(3), pp. 1695–1697 (1985).

[19] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak, "Molecular-dynamics with coupling to an external bath", *Journal of Chemical Physics* **81**(8), pp. 3684–3690 (1984).

[20] H. J. C. Berendsen, J. P. M. Postma, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath.", *Journal of Chemical Physics* **81**, pp. 3684–3690 (1984).

[21] M. Parrinello and A. Rahman, "Polymorphic transitions in single-crystals - a new molecular-dynamics method", *Journal of Applied Physics* **52**(12), pp. 7182–7190 (1981).

[22] T. E. Cheatham, P. Cieplak, and P. A. Kollman, "A modified version of the cornell et al. force field with improved sugar pucker phases and helical repeat", *Journal of Biomolecular Structure & Dynamics* **16**(4), pp. 845–862 (1999).

[23] A. D. MacKerel Jr., C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus, *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program* volume 1 of *The Encyclopedia of Computational Chemistry* , pp. 271–277, John Wiley & Sons: Chichester (1998).

[24] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids", *Journal of the American Chemical Society* **118**(45), pp. 11225–11236 (1996).

[25] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems.", *Journal Chemical Physics* **98**, pp. 10089–10092 (1993).

[26] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald potential.", *Journal of Chemical Physics* **103**, pp. 8577–8592 (1995).

[27] T. A. Halgren, "The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters", *Journal of the American Chemical Society* **114**(20), pp. 7827–7843 (1992).

[28] J. Černý, M. Kabeláč, and P. Hobza, "Double-helical – ladder structural transition in the B-DNA is induced by a loss of dispersion energy", *Journal of the American Chemical Society* **130**(47), pp. 16055–16059 (2008).

[29] D. Svozil, J. E. Šponer, I. Marchan, A. Pérez, T. E. Cheatham, F. Forti, F. J. Luque, M. Orozco, and J. Šponer, "Geometrical and electronic structure variability of the sugar – phosphate backbone in nucleic acids", *The Journal of Physical Chemistry B* **112**(27), pp. 8188–8197 (2008).

[30] A. Perez, I. Marchan, D. Svozil, J. Šponer, T. E. Cheatham III, C. A. Laughton, and M. Orozco, "Refinement of the AMBER force field for nucleic acids: Improving the description of a/g conformers", *Biophysical Journal* **92**, pp. 3817–3829 (2007).

[31] E Sobolewski, M. Makowski, C. Czaplewski, A. Liwo, S. Oldziej, and H. A. Scheraga, "Potential of mean force of hydrophobic association: Dependence on solute size", *Journal of Physical Chemistry B* **111**(36), pp. 10765–10774 (2007).

[32] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water", *Journal of Chemical Physics* **79**(2), pp. 926–935 (1983).

[33] S. Miyamoto and P. A. Kollman, "SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models", *Journal of Computational Chemistry* **13**, pp. 952–962 (1992).

[34] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, "Gaussian 03, Revision C.02", Gaussian, Inc., Wallingford, CT, 2004.

[35] Ch. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model", *The Journal of Physical Chemistry* **97**(40), pp. 10269–10280 (1993).

[36] D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K.M. Merzi, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F.Wong, F. Paesani, J. Vanicek, X.Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, and P.A. Kollman, "AMBER 10" (2008), University of California, San Francisco.

[37] J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field", *Journal of Computational Chemistry* **25**(9), pp. 1157–1174 (2004).

[38] P. Jurečka and P. Hobza, "True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine–cytosine, adenine–thymine, and their 9- and 1-methyl derivatives: Complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment", *Journal of the American Chemical Society* **125**(50), pp. 15608–15613 (2003).

[39] Ryjáček F. (2002), Ambconv.

[40] http://www.rcsb.org/pdb/home/home.do.

[41] R. E. Dickerson and H. R. Drew, "Structure of a B-DNA dodecamer .2. influence of base sequence on helix structure", *Journal of Molecular Biology* **149**(4), pp. 761–786 (1981).

[42] P. Hazel, J. Huppert, S. Balasubramanian, and S. Neidle, "Loop-length-dependent folding of g-quadruplexes", *Journal of the American Chemical Society* **126**(50), pp. 16405–16415 (2004).

[43] T. E. Cheatham, III, and P. A. Kollman, "Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution", *Journal of Molecular Biology* **259**(3), pp. 434 – 444 (1996).

[44] Chipot Ch. and Pohorrile A., *Free Energy Calculations*, Springer-Verlag (2007).

[45] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling", *Journal of Computational Physics* **23**, pp. 187–199 (1977).

[46] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, "The weighted histogram analysis method for free-energy calculations on biomolecules .1. the method", *Journal of Computational Chemistry* **13**(8), pp. 1011–1021 (1992).

[47] Kirkwood J. B., "Statistical mechanics of fluid mixtures", *Journal of Chemical Physics* **3**, pp. 300–314.

[48] B. Lee and F. M. Richards, "Interpretation of protein structures - estimation of static accessibility", *Journal of Molecular Biology* **55**(3), pp. 379–& (1971).

[49] R. E. Dickerson, M. Bansal, C. R. Calladine, Diekmann, Hunter S., W. N., and O. Kennard, "Definitions and nomenclature of nucleic acid structure parameters.", *European Molecular Biology Organisation Journal* **8** (1989).

[50] X. J. Lu and W. K. Olson, "3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures", *Nucleic Acids Research* **31**(17), pp. 5108–5121 (2003).

[51] J. Šponer, J. Leszczynski, and P. Hobza, "Structures and energies of hydrogen-bonded DNA base pairs. a nonempirical study with inclusion of electron correlation", *Journal of Physical Chemistry* **100**(5), pp. 1965–1974 (1996).

[52] P. Cieplak and P. A. Kollman, "Calculation of the free energy of association of nucleic acid bases in vacuo and water solution", *Journal of American Chemical Society* **110**(21), pp. 3734–3739 (1988).

[53] R. Sedlák, P. Jurečka, and P. Hobza, "Density functional theory-symmetry adapted perturbation treatment energy decomposition of nucleic acid base pairs taken from DNA crystal geometry", *Journal of Chemical Physics* **127**(7) (2007).

[54] J. Pranata, S. G. Wierschke, and W. L. Jorgensen, "Opls potential functions for nucleotide bases – relative association constants of hydrogen-bonded base-pairs in chloroform", *Journal of the American Chemical Society* **113**(8), pp. 2810–2819 (1991).

[55] J. L. Miller and P. A. Kollman, "Solvation free energies of the nucleic acid bases", *Journal of Physical Chemistry* **100**(20), pp. 8587–8594 (1996).

[56] R. A. Friedman and B. Honig, "A free energy analysis of nucleic acid base stacking in aqueous solution", *Biophysical Journal* **69**, pp. 1528–1535 (1995).

[57] L. C. Sowers, B. R. Shaw, and W. D. Sedwick, "Base stacking and molecular polarizability: effect of a methyl group in the 5-position of pyrimidines", *Biochemical and Biophysical Research Communications* **148**(2), pp. 790–794 (1987).

[58] K. M. Guckian, B. A. Schweitzer, R. X.-F. Ren, C. J. Sheils, D. C. Tahmassebi, and E. T. Kool, "Factors contributing to aromatic stacking in water: Evaluation in the context of dna", *Journal of American Chemical Society* **122**(10), pp. 2213–2222 (2000).

[59] O. Sinanoğlu and S. Abdulnur, "Hydrophobic stacking of the bases and the solvent denaturation of DNA", *Photochemistry and Photobiology* **3**, pp. 333–342 (1964).

[60] R. L. Hillary, S. R. Gilson, M. J. Potter, and M. K. Gilson, "The physical basis of nucleic acid base stacking in water", *Biophysical Journal* **80**, pp. 140–148 (2001).

[61] D. Chandler, "Interfaces and the driving force of hydrophobic assembly", *Nature* **437**, pp. 640–648 (2005).

[62] Naomi I. Nakano and S. J. Igarashit, "Molecular interactions of pyrimidines, purines, and some other heteroaromatic compounds in aqueous media", *Biochemistry* **9**(3), pp. 577–583 (1970).

[63] L. X. Dang and P. A. Kollman, "Molecular dynamics simulations study of the free energy of association of 9-methyladenine and 1-methylthymine bases in water", *Journal of American Chemical Society* **112**(2), pp. 503–507 (1990).

[64] E. Stofer, C. Chipot, and R. Lavery, "Free energy calculations of watson-crick base pairing in aqueous solution", *Journal of American Chemical Society* **121**(41), pp. 9503–9508 (1999).

[65] W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M Gerstein, S. C. Harvey, U. Heinemann, X. J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C. S. Tung, E. Westhof, C. Wolberger, and H. M. Berman, "A standard reference frame for the description of nucleic acid base-pair geometry", *Journal of Molecular Biology* **313**(1), pp. 229–237 (2001).

[66] D. L. Beveridge, G. Barreiro, B. K. Suzie, D. A. Case, T. E. Cheatham, S. B. Dixit, E. Giudice, F. Lankaš, R. Lavery, J.H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai, and M. A. Young, "Molecular dynamics simulations of the 136 unique tetranucleotide sequences of dna oligonucleotides. i. research design and results on d(CpG) steps", *Biophysical Journal* **87**, pp. 3799 – 3813 (2004).

[67] S. B. Dixit, D. L. Beveridge, D. A. Case, T. E. Cheatham, E. Giudice, F. Lankaš, R. Lavery, J. H. Maddocks, R. Osman, H. Sklenar, K. M. Thayer, and P. Varnai, "Molecular dynamics simulations of the 136 unique tetranucleotide sequences of dna oligonucleotides. ii: Sequence context effects on the dynamical structures of the 10 unique dinucleotide steps", *Biophysical Journal* **89**, pp. 3721 – 3740 (2005).