University of Ljubljana Faculty of Mathematics and Physics



ALL ATOM AND COARSE GRAINED DNA SIMULATION STUDIES

Julija Zavadlav

Abstract

The deoxyribonucleic acid is due to its biological importance, probably the most studied macromolecule. In this paper we focus on the computer simulation studies of DNA. First we present the two most widely used simulation methods; the Monte Carlo and the molecular dynamics method. Some tricks used in these simulations are also mentioned. The review on some applications of all atom and coarse-grained DNA simulations is given at the end.

March 2012

Contents

1	INTRODUCTION	2	
2	DNA STRUCTURE		
3	COMPUTER SIMULATION METHODS 3.1 Monte Carlo	6 6 7 8 9	
4	SOLVATION MODELS	11	
5	DNA MODELS	14	
6	SIMULATION RESULTS6.1Counterion induced attraction between DNA molecules6.2DNA in the presence of multivalent polyamine ions6.3Counterion binding6.4Sequence effects on local DNA structure	17 17 19 20 20	
7	CONCLUSION	23	

1 INTRODUCTION

The discovery of the structure of deoxyribonucleic acid (DNA) and its role in biology was one of the triumphs of 20th century science, revealing the molecular basis of genetics. To understand the mechanism of inheritance it was necessary to find the structure of DNA. The X-ray diffraction patterns of DNA done by Rosalind Franklin were certainly the important steps toward Watson and Crick's elucidation, that DNA was a right handed double helix of repeating units called nucleotides. The sequence of nucleotides stores the information required for DNA to self-assemble and maintain an organism. This is done by two key functions of DNA: the replication and transcription. The first is transmission of genetic information from parent to progeny, while the second is transformation of genetic information into a form that directs protein synthesis. In the cells and viruses DNA is found in highly organized structures, even though it is a highly charged macromolecule. On average it bears one elementary negative charge per each 0.17 nm of the double helix [1]. Because the like-charged particles should repel each other, the DNA condensation is definitely not the easiest process to understand.

It is the aim of this seminar to show, how computer simulations can contribute to our understanding of the basic features of the DNA. Computer simulations can sometimes provide us more detailed information and allow us to better understand the physical mechanisms that govern the behavior of biologically systems. They can also be seen as a bridge between theory and experiment. With use of simulations we can test the theoretical models and compare them to the experimental results. Another advantage is the possibility of simulating systems that are difficult or impossible in the laboratory, for example systems at extreme temperature or pressure [2].

Simulations of nucleic acids have shed important insights on the physics of DNA. On the atomistic level, simulations were made, for example to study the relation between the sequence and structure or the binding properties of ions, water, proteins and ligands. On the more simplified coarse-grained level a lot of simulations were conducted for the purpose of explaining the DNA condensation.

2 DNA STRUCTURE

DNA macromolecule is a long polymer also known as a polynucleotide because the double stranded DNA helix is a sequence of nucleotides, monomer units of DNA. Each nucleotide is composed of three fundamental parts: a nitrogen-rich base, sugar and phosphate group (Figure 1a). One strand of DNA if formed by successive nucleotides that are connected together via phosphate groups. As the DNA strands wind around each other, they leave gaps between each set of phosphate groups (Figure 1b). The larger gap is called the major groove, while the smaller, narrower one is called the minor groove. The two strands in double helix have a directionality. The chain ending with the terminal phosphate group linked to a fifth sugar carbon (C5') is a 5' end strand and the chain ending with a terminal hydroxyl group linked to the third sugar carbon (C3') is a 3' end strand (Figure 1b). By convention the directionality of DNA is assumed to be 5' to 3' and the individual bases are numbered according to this direction [1]. Schematically DNA could thus be represented as two directed helices running in the opposite directions.



Figure 1: Schematic structure of DNA monomer sequence. Each monomer consist of phosphate group (P), sugar (S) and base (A, T, C, G) (a). Two strands of DNA form a helical structure with a major and minor groove (b)[3].

Four nitrogen bases can be found in DNA: cytosine (C), thymine (T), guanine (G) and adenine (A). The first two belong to pyrimidine base types which have one six-member ring, while the other two belong to purines that have two fused rings, one five and one six-member ring [1]. The nitrogen bases in double-stranded DNA assemble in pairs. They are always formed by a purine and a pyrimidine. Cytosine pairs with guanine by forming three hydrogen bonds, and thymine pairs with adenine by forming two hydrogen bonds. This arrangement produces base pairs whose widths are nearly identical. The approximately uniform width is significant because any sequence can be accumulated without much alteration in overall structure. This complementary base pairing is the key property that allows DNA to function as the mechanism of information storage and inheritance [1].

The global structure of DNA double helix can be described with the following helical parameters;

- helix sense refers to the handedness of the double helix,
- helix diameter is the geometric diameter of the helical cylinder,
- helix pitch measures the distance along the helix axis for one complete turn,
- number of residues per turn is the number of base pairs for every complete helix turn.

To this day three different forms of DNA were characterized and example structures are shown in Figure 2. Under conditions relevant for the life of cells, the dominant structure is the so called B form. It is a right-handed helix with diameter of 20Å, helix pitch of 34Å and has 10 nucleotides per turn. The form termed A-DNA emerged from studies of nucleic acid fibers at much lower value of relative humidity. Like the previous form, the A-DNA is also right-handed, but whereas in B-DNA the base pairs lie almost perpendicular to the helix axis in A-DNA the base pairs are significantly tilted with respect to the helix axis. A rather surprising finding was a peculiar left-handed helix which was observed at high salt concentrations and named Z-DNA for its zigzag pattern. Because

of the increased salt concentration the electrostatic repulsion between the two backbones is weakened due to the increased shielding. The resulting structure has a smaller 18Å helix diameter and larger pitch of about 45Å. In contrast to the A and B-DNA models, specific sequences are required for Z-DNA (alternating GC). Though all three forms are biologically relevant, the B-DNA is by far the most studied one.



Figure 2: The structure of different DNA helix forms; the B-DNA (a), A-DNA (b) and Z-DNA (c).

The various fiber diffraction studies have shown that the base pairs arrangement within each DNA form can vary quite a lot. The base pairs thus have different degrees of freedom that can be quantified by translational and rotational deformations of base pairs. The common nomenclature for parameters used to describe the geometry of nucleic acid helices was defined at the EMBO workshop in Cambridge [4]. This parameters are described in Table 1.

Local variables that relate the local coordinate frames of two					
successive base pairs.					
Twist (ω)	Roll (ρ)	Tilt (τ)			
	- Sector	-			
Rise (D_z)	Slide (D_y)	Shift (D_x)			



Table 1: Definitions of various parameters used to describe the geometry of nucleic acid helices in a coordinate frame, where the x direction of local base pair coordinate is pointing along the short axis of the base pair, the y direction along the long axis and the z direction parallel with the helix axis [4].

For DNA molecules in solution the hydration effects are of great importance, since water plays a central role in stabilizing a particular helical conformation [1]. DNA molecule is neither hydrophobic nor hydrophilic. It belongs to a special intermediate group of amphiphilic molecules that share properties from both classes: the DNA backbone is hydrophilic and the bases are hydrophobic. This frustrated, amphiphilic character of DNA together with the flexibility of the backbone, produces the famous double-helical structure of DNA. The distance between adjacent sugars along the phosphate backbone in DNA chain is about 6Å [5]. It is determined by the structural properties of phosphate and sugar and cannot vary much. On the other hand the bases are not soluble in water and tend to avoid water molecules as much as possible. The bases that attract each due to the hydrophobic force tend to be at a separation of 3.4Å. We can see how a structure of a straight ladder would not satisfy the phosphate-phosphate bond length as well as the stacking separation between bases and how gradually twisting the ladder into the shape of helix could. Because of the twist of the base pairs around the axis of the molecule part of the bases still get exposed to water. Due to this fact, the base-pairs reorient (with twist and propeller-twist) in such a way that they minimize the

extent of exposure to water [5].

In order for the information stored in DNA sequences to be useful, the bases must be accessible to enzymes. It is therefore necessary that the interactions between the bases is only marginally stable. Since hydrogen bond is a relatively weak bond with the characteristic energy scale of the order of thermal energy k_BT , the helix can be opened and the sequence read [1]. This week hydrogen bond characteristic is also the reason why at high temperatures (above $70 - 80^{\circ}$ C) the two strands of DNA fall apart. In other words, DNA melts.

3 COMPUTER SIMULATION METHODS

The two main families of simulation technique are molecular dynamics (MD) and Monte Carlo (MC). Additionally, there is a whole range of hybrid techniques which combine features from both.

3.1 Monte Carlo

Monte Carlo is a stochastic computer simulation method that uses probabilistic method to solve problems [1]. In the Metropolis algorithm a trail state x' is generated by a random perturbation of initial state x. Trial state is accepted if the corresponding new energy is lower. If, however, E(x') > E(x), then the new state is accepted if the probability

$$p = exp(-\beta \Delta E) \tag{1}$$

is grater than a uniformly generated random number on interval (0,1). In Eq. (1) $\Delta E = E(x') - E(x) > 0$ and $\beta = 1/k_BT$. Consequently, the sequence tends to regions with low energies. But since the states with higher energies have a nonzero probability of acceptance, the method can overcome barriers in conformational space and escape from local minimum [1].

An extension of the Metropolis algorithm is often employed as a global minimization technique known as simulated annealing, where the temperature is lowered as the simulation evolves.

The type of state perturbation depends on the system. It can be translational, rotational, local or global move [1]. Specifying appropriate MC move can sometimes be an art by itself. At first glance MC move would be perturbation of all particles independently. However in biomolecular simulations, moving all particles is highly inefficient, since it leads to a large percentage of rejections. It is thus more common to perturb one or few particles at each step. The unwritten rule is to aim for a perturbation that yields about 50% acceptance.

The simplicity and general applicability of Monte Carlo approach has long been exploited for molecular applications. One advantage is that it can be easily applied to discontinuous potentials, like the square well potential often used for colloidal suspensions. It is also very suitable for a description of polyelectrolyte systems within the frame of continuum solvent models. However, for models with explicit solvation, the MC method is not so efficient. The main reason is that in the liquid state, the molecules are closely packed and the fraction of accepted MC moves becomes too small.

3.2 Molecular dynamics

The molecular dynamics approach is very simple in principle. We simulate the motion of a system through step by step calculation of particles coordinates and velocities according to Newtons equation of motion [6]. By following the dynamics of a system in space and time, we can obtain a

rich amount of information concerning the structural and dynamic properties. For a system composed of N particles with coordinates $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$ and momentum $\mathbf{p}^N = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_N)$ the classical equations of motion are

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} \ and \ \dot{\mathbf{p}}_i = \mathbf{f}_i$$

$$\tag{2}$$

Many methods exist to solve this system of coupled ordinary differential equations by performing numerical integration of them. Here we introduce one of such method called velocity Verlet algorithm [6]. This frequently used method can be illustrated by the following pseudo-code:

$$dostep: 1, nstep$$

$$p = p + 0.5 * f * dt$$

$$r = r + p * dt/m$$

$$f = force(r)$$

$$p = p + 0.5 * f * dt$$

$$enddo$$

Classical MD simulations are unsuitable for low temperatures. There the energy gaps among the discrete levels of energy, dictated by quantum physics, are much larger than the thermal energy [1]. This discrete description of energy states becomes less important as the temperature is increased or the frequencies associated with motion are decreased. Roughly classical behavior is approached for frequency/temperature combinations for which

$$\frac{h\nu}{k_BT} \ll 1,\tag{3}$$

where h is Plancks constant, ν the frequency, k_B Boltzmanns constant, and T the temperature [1]. The highest requency mode in system determines the timestep. In standard explicit schemes the timesptep is about 1 fs [1]. This implies that one million steps have to be made to cover only a nanosecond. The limits MD for simulating slow molecular events. It is also a computationally expensive method, but only MD can ultimately yield detailed dynamic information such as time correlation functions, diffusion, and other transport properties.

3.3 Force Fields

What ever the simulation method we use, to simulate a system we have to assume some kind of model i.e. the interactions between the constituting parts of our system. For example in order to use MD simulation method we need to be able to calculate forces \mathbf{f}_i acting on particles. Using the relation

$$\mathbf{f}_i = -\frac{\partial}{\partial \mathbf{r}_i} U(\mathbf{r}^N) \tag{4}$$

we can derive the forces from a potential energy $U(\mathbf{r}^N)$, where $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$ represents the complete set of 3N coordinates. The form and parameters of mathematical functions used to describe the potential energy of a particular system is in molecular simulations called a force field. Many of the atomistic force fields in use today for molecular systems can be interpreted in terms of a relatively simple four component potential composed of the intra- and inter-molecular interactions [2]. One such functional form of a force field is:

$$U(\mathbf{r}^{N}) = \frac{1}{2} \sum_{bonds}^{bonds} k_{b}(r_{ij} - r_{0})^{2} + \frac{1}{2} \sum_{angles}^{bonds} k_{\theta}(\theta_{ijk} - \theta_{0})^{2} + \frac{1}{2} \sum_{angles}^{angles} k_{\phi}(1 + \cos(m\phi_{ijkl} - \gamma)) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\epsilon_{ij} \left\{ \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right\} + \frac{q_{i}q_{j}}{4\pi\epsilon_{0}r_{ij}} \right)$$

$$(5)$$

The first term in Eq. (5) describes the stretching of bonds, the second term the opening and closing of angles and the third term is a torsional potential that models how the energy changes as a bond rotates. As we can see this bonded interactions can be described with simple functions based on Hooke's law where we assign certain energetic penalty when the bonds and angles deviate away from their equilibrium values [2]. The non-bonded contribution is described with the forth term and modeled using a Coulomb potential for electrostatic interactions and a Lennard-Jones potential for van der Waals interactions. This term is calculated between all pairs of atoms that belong to different molecules.

To define a force field one must specify not only the functional form, but also the parameters such as k_b in Eq. (5). These parameters are set to reproduce some experimental results, typically the structural properties of a system. Some information about them can also be obtained from quantum mechanical calculations. Important point is that force fields are empirical, so there is no correct form for a force field [1]. As a consequence, simulation results will dependent on the force field used.

3.4 Non-bonded interactions

Computing the non-bonded contribution to the potential turns out to be quite challenging. The non-bonded interactions have a nonzero value even for large inter-particle distances. This implies a huge number of pairwise calculations and in practice we can not compute them all. For a system involving N particles combinatorics tells us that the number of distinct particle pairs is equal to N(N-1)/2. The simplest approximation is achieved with a potential cutoff distance, beyond which the interaction is small enough to be neglected. Even with cutoffs checking whether or not two particles are within a cutoff distance is still very time consuming.

Improving the speed of a program can be achieved with Verlet list [6]. Figure 3a shows the schematic representation. This list contains the neighbors of each atom with pair separation smaller than the value of r_{list} (dashed line on Figure 3a), which is determined by the sum of r_{cutoff} value (solid line on Figure 3a) and the r_{skin} , the so-called skin. In this way only the pairs on the list are considered in the computation of non-bonded interactions. The list has to be reconstructed before any unlisted pairs come within interaction range. Reconstruction is of coarse more frequent for smaller lists with smaller r_{skin} , but smaller lists also save more CPU time.

For larger systems another technique becomes preferable where simulation box is divided into cells [6]. Example is shown for cubic box on Figure 3b. These cells are chosen so that the side of the cell is greater than the potential cutoff distance. For each cell a list keeps track of the particles in the cell. Non-bonded interactions are computed between the atom of interest (white circle on Figure 3b) and the atoms in the same cell and in nearest neighbour cells (gray area on Figure 3b).



Figure 3: Schematic representation of Verlet list construction (a). The potential cutoff range r_{cutoff} is shown with a solid circle, while the list range r_{list} is shown with dashed circle. The list must be reconstructed before particles originally outside the list range (black) have penetrated the potential cutoff sphere. The cell structure (b). In searching for neighbours of an atom, it is only necessary to examine the atom's own cell, and its nearest-neighbour cells (collored in gray) [6].

Compared to Verlet list a certain amount of unnecessary work is done because the search region is cubic, not spherical [6].

3.5 Periodic boundary conditions

With computer simulations we want to sample finite systems with small number of particles. At the same time these particles have to experience the same forces as if they were in bulk fluid. This means that unless surface effects are of particular interest, periodic boundary conditions (PBC) have to be used [6]. Here the edge effects are minimized by surrounding the simulation box with translated image copies of itself. The illustration for two dimensional box is presented in Figure 4. During the simulation only the coordinates of particles in the real cell are recorded and propagated, because the image coordinates can be computed simply by adding or subtracting the box side. If a particle leaves the box during the simulation then it is replaced by an image particle that enters from the opposite side, as illustrated in Figure 4. The total number of particles within the simulation box thus remains constant.

Particle interacts with both the real particles and the imaginary ones are (see Figure 4). Usually minimum image convention is adopted where interaction between two particles is computed only once i.e. the one with the smallest inter-particle distance [6]. To accomplish this the dimension of the box side has to be at least twice as big as the cut-off radius used for non-bonded interactions [6]. Also when a macromolecule, such as a DNA, is studied, this restriction alone is not sufficient, because a single solvent molecule should not be able to see both sides of the macromolecule [1]. Although PBC are widely used in computer simulations, they have some drawbacks because of the imposed artificial periodicity. This effects can be evaluated empirically by comparing the results obtained with a variety of cell shapes and sizes.

The cubic cell is the simplest periodic system to visualize and to program. However, any cell shape that fills entire three dimensional space by translation operations can be used. Shapes that satisfy this condition are: the cube, hexagonal prism, the truncated octahedron and rhombic dodecahedron

Figure 4: Periodic boundary condition. As a particle moves out of the simulation box, an image particle moves in to replace it. In calculating particle interactions within the cutoff range, both real and image neighbours are included [6].

[1]. The use of a shape other than cubic may be particularly convenient for simulations of a macromolecule surrounded by solvent molecules. In such systems, it is usually the behavior of the central macromolecule that is of most interest. Therefore it is desirable to spend as little of computer time as possible simulating the solvent far away. At same periodic distance the volume of rhombic dodecahedron is only 71% of the volume of cube, which means that fewer solvent molecules are required to fill the box [1]. For octahedron the corresponding volume is 77% of a cube. It is also sensible to choose a periodic cell that reflects the underlying geometry of the system. The rhombic dodecahedron and the truncated octahedron are almost spherical in shape. They are therefore better suited to the study of spherical macromolecules like proteins [1]. For cylindrical molecules like DNA a hexagonal prism is more appropriated.

3.6 Constrained Dynamics

As we already mentioned the applicability of molecular dynamics is seriously limited by the necessary use of small timesteps and thus the simulation lengths that can be achieved with current computer capabilities. The fastest motions in the system that determine the timestep (~ 1fs) are bond stretching terms, in particular the bonds involving hydrogens [1]. A vary practical solution to this problem is provided by constrained dynamics, where bond lengths are constrained to have fixed length. In this way bonds are made rigid, the highest frequency motions are removed and the timesteps can be lengthened (~ 5fs) [1]. The bond length constraint can be expressed as

$$\chi = \mathbf{r}_{ij}^2 - r_0^2 = 0, \tag{6}$$

where $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ is the vector associated with the rigid bond between atoms i and j and r_0 the constrained or equilibrium bond value [6]. In classical mechanics, constraints are introduced

Figure 5: Examples of periodic domains in 3D used for biomolecular simulations; an orthorombic box containing a solvated protein (a), a hexagonal prism containing a solvated DNA dodecamer (b) and truncated octahedron containing a solvated protein [1].

through the Lagrangian formalism, where the equation of motion is written as

$$m_i \ddot{\mathbf{r}}_i = \mathbf{f}_i + \Lambda \mathbf{g}_i \ , \ \mathbf{g}_i = -\frac{\partial \chi}{\partial \mathbf{r}_i},$$
 (7)

where Λ is the undetermined Lagrange multiplier. The SHAKE algorithm incorporated in the original Verlet algorithm was the first algorithm developed to satisfy bond geometry constraints. In this scheme for each time step the equations of motion are solved in the absence of constraints, than the magnitudes of the constraint forces are determined and the particles positions are corrected. It aniterative method, where each constraint is satisfied in turn until convergence is reached. Angle constraints can also be accommodated by recognizing that an angle constraint simply corresponds to an additional distance constraint. For example, rigid water molecules can be constrained with the use of three distance constraints.

When velocities appear in the integration algorithm (velocity Verlet algorithm) these must be corrected as well as the positions [6]. The time derivative of the Eq. (7) gives constraints on the velocities

$$2\mathbf{r}_{ij}\mathbf{r}_{ij} = 0,\tag{8}$$

where $\mathbf{v}_{ij} = \mathbf{v}_j - \mathbf{v}_i$ is the relative velocity acting on the atoms i and j constituting the rigid bond. In geometrical sense this means that the net velocity acting on the bond must be perpendicular to that bond. Velocity correction is performed by the iterative RATTLE module. The iterative methods can be quite time consuming, so for rigid water molecules an analytical solution of SHAKE and RATTLE, the so-called SETTLE algorithm is preferred.

4 SOLVATION MODELS

To obtain meaningful simulations results and to be able to compare them with experiments it is necessary to simulate DNA in its natural environment. This means that we have to properly incorporate the surrounding aqueous ionic solution into our model. There are two different methods of solvent representation; the implicit and explicit solvation. Implicit solvation is a method where solvent is modeled as a continuous medium. Water molecules are not actually present in our system, but rather their presence is accounted for in an average sense through the form of electrostatic potential [1].

Let us consider the solution of macro-ions, couterions and added salt. Then the charge density of the marco-ions and the mobile ions

$$\rho(\mathbf{r}) = \rho_{macro}(\mathbf{r}) + \sum_{i} z_{i} e n_{i}(\mathbf{r})$$
(9)

can be related to local electrostatic potential Φ through Poisson's equation

$$\Delta \Phi(\mathbf{r}) = -\frac{4\pi}{\epsilon} \rho(\mathbf{r}),\tag{10}$$

where ϵ is the dielectric constant of the continuum (aqueous medium) in which the ions are dissolved [7]. For point charges this equation reduces to Coulombs law. The minimization of system's free energy with respect to ion number density leads to the condition that the ions obey the Boltzmann distribution, where

$$n = n_0 e^{-ze\Phi/k_B T}.$$
(11)

To derive this equation one has to make the so-called mean-field approximation, where the correlations between the ion positions are neglected. Combining the Eq. (10) with Eq. (11) yields the well known Poisson-Boltzmann (PB) equation:

$$\Delta \Phi = -\frac{4\pi e n_0}{\epsilon} \sum_i z_i e^{-e_i \Phi/k_B T}.$$
(12)

The PB equation is the basis for modern electrolyte theory [7]. It is a non-linear equation and in general the analytic solutions are not available. In cases where energies are much smaller than the thermal energy, i.e. when $e\Phi/k_BT \ll 1$, the PB equation can be linearized by Taylor-series expansion of the exponent. The result is the Debye-Hückel approximation

$$\Delta \Phi(r) = \kappa^2 \Phi(r), \tag{13}$$

where

$$\kappa^{-1} = \frac{z^2 e^2 n_0}{\epsilon_0 \epsilon k_B T} \tag{14}$$

is the Debye screening length. The solution is known as Yukawa potential

$$\Phi(r) = B(\kappa) \frac{-\kappa r}{\epsilon r}.$$
(15)

Thus, Debye-Hückel theory produces an effective electrostatic potential in which the Coulombis interactions are screened by ions [7]. For physiological ionic strengths, such as 0.15M, the Debye length is approximately 8Å [1]. Figure 6 compares the Coulombic potential (1/r) to the screened Coulombic potential $(exp(\kappa r)/r)$ for two values of κ corresponding to monovalent salt concentrations of 0.15M and 0.015M.

Because of the reduced number of degrees of freedom, implicit solvation approach can be quite effective, especially when combined with coarse-grained molecular models. The drawback of such models are their incapability to properly describe solvation around macromolecules where dielectric

Figure 6: The Coulomb potential (1/r) is compared to the screened Coulomb potential as a function of distance r for two values of κ that correspond to 0.15M and 0.015M monovalent salt concentrations and Debye lengths of about 8 and 25Å [1].

Figure 7: Classification of atomistic water models according to the number of interaction sites (a). The SPC water model (b). predominance of positive charge is on H atom (white) and excess negative charge on O atom (red). The model assumes an ideal tetrahedral shape with the HOH angle equal to 109.47° , instead of the observed value of 104.5° so that the distance between hydrogens is equal to 1.63\AA .

constant is smaller than in the bulk [8].

This problem is avoided with the explicit solvation, where the water molecule and ions are explicitly present in the system. Many different atomistic water models have been proposed. They can be classified by the number of interactions points used (atoms plus dummy sites) or whether the structure is rigid or flexible.

The three-site models have three interaction sites, corresponding to the three atoms of the water molecule. In the case of a Simple Point Charge (SPC) model there is a predominance of positive charge of $0.41e_0$ on the H atoms and excess negative charge of $-0.81e_0$ on the O atom. The equilibrium distance between the O and H is 1.0Å. The model assumes an ideal tetrahedral shape with the HOH angle equal to 109.47° , instead of the observed value of 104.5° . If the model is rigid than the only interactions are the non-bonded ones. The electrostatic is modeled using Coulomb's law and the van der Waals interactions using Lennard-Jones potential. If however the water model is flexible, than an additional bonded interactions are included. Other well known examples of the three-site models are the TIP3P and SPC/E water models.

To improve the electrostatic distribution around the water molecule the 4-site models can be used, which place the negative charge on a dummy atom (labeled M on the Figure 7) placed near the O atom along the bisector of the HOH angle. Such an example is the TIP4P water model. Even more advanced but in practice rarely used models are the 5-site models like TIP5P that place the negative charge on dummy atoms (labeled L on the Figure 7) representing the ione pairs of the oxygen atom.

While such elaborate atomistic solvation models can yield invaluable information on the rich complexity of biomolecular environments, they are clearly computationally expensive. In this regard coarse-grained solvent models can be advantageous. One such model, the mW-ion model, represents each water molecule and an ion atom (Na⁺ and Cl⁻) by a single particle [9]. Since all the particles including the ions are chargeless, there is no electrostatics present in the system and the particles interact with each other through a very short-ranged potentials. The hypothesis behind the model is that even though Coulombic forces act over a long distance, they are shielded by the net effect of the surrounding charges and waters partial charges.

In typical atomistic water models hydrogen bonding and associated tetrahedral water arrangement is produced by competition of electrostatic attraction and short-ranged repulsion. In other words, short-ranged ordering is effected through long-ranged interactions. In the case of mW-ion model the short ranged ordering is produced with the use of short-ranged interactions of the Stillinger-Weber (SW) potential [9]. This potential uses of a combination of two and three-body potentials that encourage tetrahedral configurations of the water molecules and has the following functional form

$$U(\mathbf{r}^{N}) = \sum_{i} \sum_{j>i} A\epsilon_{ij} \left\{ B\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{p} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{q} \right\} e^{\frac{\sigma_{ij}}{r_{ij} - a_{ij}\sigma_{ij}}} + \sum_{i} \sum_{j\neq i} \sum_{k>j} \lambda_{ijk} \epsilon_{ijk} [\cos\theta_{ijk} - \cos\theta_{0}]^{2} e^{\frac{\gamma\sigma_{ij}}{r_{ij} - a_{ij}\sigma_{ij}}} e^{\frac{\gamma\sigma_{ik}}{r_{ik} - a_{ik}\sigma_{ik}}}$$
(16)

In Eq. (16) the r_{ij} is the distance between particles i and j and θ_{ijk} is the angle between j-i-k particles. The large number of parameters of this force field allow customization of the interaction behavior to reproduce the experimentally observed properties of liquid water [9]. The energy scale ϵ influences the depth of the interaction, while σ determines the particle size. The parameter λ tunes the strength of the tetrahedral interactions by applying an energy penalty to configurations in which the three involved particles form (16) an angle other than the specified tetrahedral angle. The differences in solvation of Na and Cl ions, i.e. the ability of Na⁺ to disrupt the native tetrahedral arrangement and of Cl⁻ to integrate within this organization, can be incorporated with appropriate λ . To model the repulsion between NaNa and ClCl pairs additional shielded Coulomb Yukawa potential is used. The values of r_{cutoff} for Yukawa potential does not exceed 7Å, while the lack of hydrogen atoms, the simulations with mW-ion model are 2 orders of magnitude faster than rigid atomistic water.

5 DNA MODELS

Several models of DNA are available in the literature. These range from fully atomistic representations, in which all atoms including the solvents are considered explicitly, to highly coarse grained, where collections of several hundred atoms are represented as a single particle. While atomistic description would at first glance appear desirable, as more degrees of freedom are included in a model the computational requirements increase significantly, thereby restricting severely the simulation length and size. Ultimately the best choice of DNA model will depend on the processes we aim to study. The challenge is to include just enough detail in a model to capture the relevant physics behind it.

Most of early DNA simulations were coarse-grained models. The primary objective of these studies was to evaluate the applicability of analytical theories describing electrostatic interactions between counterions and DNA [8]. The so-called **primitive model** is the simplest approach, in which a single DNA is treated as a uniformly charged hard cylinder with radius of about 12Å. The mobile ions are considered as point charges or small rigid charged hard spheres [10].

Figure 8: The primitive (a) and grooved (b) coarse-grained model for DNA.

More elaborate DNA models such as the **grooved model** include specific details of DNA's structure. Within this model DNA is simulated as an infinitely long hard cylinder with charge located on the sites corresponding to phosphate group in B-form of DNA. This model thus in an idealized way incorporates effects of discrete charge localization on the DNA surface and the major and minor groove structure of the double helix [10]. In both the primitive and the grooved model the solvent is usually described implicitly [8]. If the phosphate group charges and the ions are modeled as point charges with repulsive r^{-12} potential of effective diameter σ , than the total potential energy of the system is

$$U = \sum_{i} \frac{qq}{4\pi\epsilon_0\epsilon r} + \sum_{i} \left(\frac{b}{\sigma}\right)^{12} , \ b = \frac{(\sigma_i + \sigma_j)k_BT}{2}.$$
 (17)

Coarse-grained models are by definition unable to deal with sequence specific processes and structural properties of DNA. At the other end of spectrum atomistic models provide the highest degree of detail. Studies employing these representations usually involve simulation of small DNA oligomers approximately ten base pairs in length, while the simulation times are on the order of 10ns. At present, the most often used force fields are AMBER, CHARMM and GROMOS. It was found that these force fields may give somewhat different DNA structures [1]. However, the hydration structure and ion distribution are very similar. This is because the force fields differ mainly by parameters describing intramolecular DNA interactions, whereas parameters describing forces between water, ions, and DNA are almost the same.

For many relevant phenomena or processes of interest, however, the approaches mentioned above, either atomistic or highly coarse grained, are inadequate. Atomistic simulations are still intractable for long simulations and typical CG implementations are too simplified to provide the necessarily resolution or molecular detail required. To this end mesoscale models with different levels of coarsening have been proposed that focus on DNA's persistence length, hybridization, denaturation and melting [11].

A mesoscale 3SPN-DNA model parameterized by Knotts reduces the complexity of a nucleotide to three interactions sites which correspond to phosphate, sugar, and base [11]. Additionally the use of four different base sites allows differentiation between types of bases present in DNA.

Figure 9: A mesoscale 3SPN-DNA model of DNA. Grouping of the atoms for each coarse grain site (a). Single-strand topology (b). Model of a 13 base pair oligonucleotide (c) [11].

For a cytosine nucleotide the grouping of atoms into three sites is depicted in Figure 9a. The backbone phosphate and sugar sites are placed at the center of mass of the respective moiety. For purine the interaction site is placed at the N1 atom position and for pyrimidine bases at the N3 atom position [11]. Figure 9b shows the topology of a single strand, while Figure 9c shows a 13 base-pairs that illustrates how the model captures the characteristic major and minor grooves of DNA. The potential energy of the presented system includes six distinct contributions

$$U_{total} = U_{bond} + U_{angle} + U_{dihedral} + U_{stack} + U_{bp} + U_{ex};$$

$$(18)$$

$$U_{bond} = \sum_{i}^{N_{bond}} [k_1(d_i - d_0)^2 + [k_2(d_i - d_0)^4], \quad U_{stack} = \sum_{i < j}^{N_{st}} 4\epsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \tag{19}$$

$$U_{angle} = \frac{1}{2} \sum_{i}^{N_{angle}} k_{\theta} (\theta_i - \theta_0)^2, \qquad \qquad U_{bp} = \sum_{bp}^{N_{bp}} 4\epsilon_{bp_i} \left[5 \left(\frac{\sigma_{bp_i}}{r_{ij}} \right)^{12} - 6 \left(\frac{\sigma_{bp_i}}{r_{ij}} \right)^{10} \right], \quad (20)$$

$$U_{dihedral} = \sum_{i}^{N_{dihedral}} k_{\phi} [1 - \cos(\phi_i - \phi_0)], \qquad U_{ex} = \sum_{i < j}^{N_{ex}} 4\epsilon \left[\left(\frac{\sigma_0}{r_{ij}}\right)^{12} - \left(\frac{\sigma_0}{r_{ij}}\right)^6 \right] + \epsilon.$$
(21)

The terms U_{bond} , U_{angle} and $U_{dihedral}$ in Eq. (18) are typical expressions for intramolecular bond stretching, opening and closing of angles and the bond rotation. The remaining terms of Eq. (18) describe various pairwise, nonbonded interactions. The base-pair stacking and backbone rigidity is modeled with an intra- strand term U_{stack} term. The term U_{bp} describes hydrogen bonding between any complementary base pair and acts both intra and inter-strand, while the U_{ex} term describes excluded volume interactions. Contributions for U_{stack} , U_{bp} , and U_{ex} are mutually exclusive, meaning that a pair of sites contributes to one and only one of these terms. For the implementation with implicit solvation additional Coulombic interactions for phosphate sites are taken into account using the Debye-Hückel approximation [11]. Since Debye length is related to the ionic strength of the solution, different salt concentrations are modeled by adjusting the Debye length to appropriate value. Alternatively this interaction is omitted for the implementation with explicit solvation [12]. DeMille et.al have combined the 3SPN model with coarse-grained mW-ion model and named it mW/3SPN-DNA model [12].

6 SIMULATION RESULTS

6.1 Counterion induced attraction between DNA molecules

It is well documented that an addition of multivalent ions may lead to condensation of DNA into an ordered phase [13]. Parsegian and co-workers have studied the DNA-DNA interactions in condensed gels of ordered DNA by means of an osmotic stress technique [14]. In this measurements DNA molecules in salt solution is mixed with a water soluble polymer (usually polyethyleneglycol - PEG). The result of mixing is hexagonally ordered DNA in equilibrium with the polymer phase. The osmotic pressure exerted on DNA can be varied by varying the polymer concentration, leading to equilibrium at various DNA-DNA distances, measured by X-ray diffraction [14]. In this way, the osmotic pressure-distance curves can be obtained at different salt concentrations.

These force measurements, performed on DNA in univalent salt solutions, showed repulsive interactions between DNA double helices. For some multivalent couterions however, effective attraction between DNA molecules has been seen [7]. This may be taken as an indication, that a physical mechanism is involved that is connected to the electrostatic interaction between the counterions and DNA. Usually, the electrostatic interaction between highly charged polyelectrolytes is treated within the framework of mean-field Poisson-Boltzmann equation [7]. For any ion solution composition, this theory gives an effective repulsive force between like-charged macromolecules. The only effect of ionic composition is on the magnitude of the force. This theory is thus not capable of explaining the condensation to an ordered DNA phase induced by multivalent ions.

Figure 10: Monte Carlo simulation calculation of osmotic pressure in an ordered DNA system in equilibrium with a 2:1 electrolyte bulk phase of varying concentration. The filled squares are experimental results for a DNA ordered gel phase in equilibrium with a 0.05 M MnCl₂ bulk solution [10].

The grand canonical Monte Carlo simulations of the grooved DNA model revealed better agreement

with the mentioned experimental results [10]. The osmotic pressure of an ordered DNA phase in electrolyte solution was calculated by determining free energy differences for different DNA-DNA separations. The result for a varying concentration of divalent salt is shown on Figure 10. As we can see for larger salt concentrations there is a negative pressure region at DNA-DNA distances between 26-32Å. This region of attraction coincides with the experimental osmotic pressure data obtained for DNA in equilibrium with a MnCl₂ solution is added [10]. It is clear that the higher concentration of divalent salt leads to a stronger attraction.

Essentially the same result was obtained by Grønbech-Jensen et al [15]. They performed Brownian dynamics simulations of two parallel uniformly charged rods with counterions, but without added salt. Allahyarov and Lõwen did molecular dynamics simulations of grooved DNA model, but with added salt [16]. Again, the general conclusion of these works was that inclusion of ion correlations, that are neglected in the mean-field PB theory, gives an attractive contribution to the electrostatic force which in the case of multivalent ions leads to the net attractive force. The effect of ion correlations can be explained as follows. If a counterion is present at some point near the DNA's surface, it will decrease the probability for other counterions to be around it. The decrease in local counterion density causes an effective attractive force that draws the counterions closer to the DNA's surface [8].

The fact that the PB approximation underestimates the ion density in the nearest layer next to the DNA's surface can be seen form ion density distribution plots around DNA or from integrated charge curve, which is defined as ion distribution integrated from the surface to some distance r [8]. It represents the total counterion charge (per DNA length) within the cylinder of radius r around the DNA. The integral charge curve is shown in Figure 11. for the mixture of monovalent and divalent counterions with monovalent coions. Comparison of the PB and MC simulation results shows that the amount of divalent ions within 5Å from the DNA surface in the PB approximation is underestimated by about 20% [8].

Figure 11: Radially integrated counterionic $(q^{2+} \text{ and } q^{+})$ and total (q) charge for a 0.155 M NaCl and 0.022 M MgCl₂ salt mixture around a cylindrical model of DNA. Points are MC simulation result, while the solid lines correspond to PB theory [8].

The interaction of mobile ions with DNA depends strongly on the ion charge. Counterions of higher valency are more strongly attracted to DNA than monovalent ions and as a result they force

low-valency ions out from the nearest vicinity of the polyelectrolyte [8]. It has been experimentally demonstrated that the size of the ions additionally plays an important role. The interaction is always repulsive for solutions with Mg^{2+} ions, while solutions containing ions with smaller hydrated radius like Mn^{2+} , have been found to cause ordered DNA phase [10].

Figure 12: Osmotic pressure in the ordered DNA system with +2 counterions for various ion diameter sigma = $0\text{\AA}(1)$, $1\text{\AA}(2)$, $4\text{\AA}(3)$, $5\text{\AA}(4)$ and $6\text{\AA}(5)$ [10].

In this regard osmotic pressure was calculated for a salt free solution with divalent counterions of different diameter size [10]. It can be seen in Figure 12 that the ion size has a crucial influence on the pressure for distances less than 30Å. For point charges and small ion diameters that correspond to naked ions without hydrated shell there is a strong attractive force at small DNA separation. For a typical size of hydrated ion of 4Å the osmotic pressure curve is similar to experimental osmotic pressure in 0.05 M MgCl₂ solution, where the attraction region is for distances between 27 and 33Å. Further increasing the ion size leads to disappearance of the attractive region. This osmotic pressure-distance curves are similar to those obtained with DNA in MgCl₂ solutions [10]. Thus a small change of the ion size may cause a transition from a purely repulsive interaction to an attractive one.

6.2 DNA in the presence of multivalent polyamine ions

Studies of DNA interacting with complex multivalent molecular ions, such as spermidine and spermine, are appealing because of their role in living systems. There is no simple analytical theory that describes DNA in the presence of such counterions, so computer simulations are the only way of making theoretical investigations of such systems [8].

In Ref.[17], the spermidine (Spd^{3+}) counterions were considered in the frame of the primitive and grooved DNA model as a flexible chain of three monovalent ions connected by harmonic bonds. It was found that because of the effects of the nonlocal charge distribution and internal degrees of freedom, the binding affinity of spermidine molecule is reduced compared with that of a simple trivalent (metal) ion. The binding affinity is even below the predictions of the PB theory [17]. The reason for this behavior is that the interaction between a counterion with three distributed charges and a polyelectrolyte is weaker than that of a three-valent point charge counterion. In addition, a

spermidine molecule bound to a polyion has less space to move in and, hence, lower entropy as the same molecule in the bulk. As a consequence divalent counterions can compete with palyamines for association to DNA. This can be seen on Figure 13, where density distributions are shown for DNA in NaCl solution containing Spd^{3+} and Mg^{2+} counterions.

Figure 13: Density distribution of Spd^{3+} (0.02 M), Mg^{2+} (0.1M) and Cl^- in different models; the Poisson-Boltzmann (PB), primitive model (HC) and the grooved model (GD) [17].

6.3 Counterion binding

The question of sequence-specific counterion binding to DNA has a principal importance to our understanding of mechanisms of DNA recognition. A series of studies was done in order to elucidate the binding sites of different counterions. It was found that monovalent ions interact with DNA in a very different manner [8]. Li⁺ ions bind almost exclusively to the phosphate groups of DNA. Na+ ions bind prevailingly to bases in the minor groove, while Cs^+ ions bind directly to sugar in minor groove. Flexible polyamine molecules have several binding modes, interacting with different sites in an irregular manner. That is why polyamine molecules are not seen in X-ray diffractions of DNA [18]. Some studies have shown that spermidine and spermine binding is prevalent in minor groove, while there seem to be no strong binding sites in the major groove [18].

6.4 Sequence effects on local DNA structure

It is well recognized that base sequence exerts a significant influence on the properties of DNA and plays a significant role in protein - DNA interactions vital for cellular processes. The geometric, electrostatic and mechanical differences between the base-pairs affect the local DNA structure. Such local variations translate into global structural effects and in turn, affect key functional processes, from replication and transcription to genome packaging and processing.

To understand and predict sequence-induced effects one requires a detailed knowledge of base sequence effects on local properties of DNA. Since crystallography and NMR spectroscopy do not provide enough high-resolution data to make reliable sequence effects predictions, the use of molecular simulations for this purpose is especially attractive. A large-scale atomistic systematic study to analyze and understand the impact of base sequence on structure and structural fluctuations, has been carried out by the Ascona B-DNA consortium (ABC), a collaboration of various groups established in Switzerland in 2002 [19].

Their initial assumption was that the sequence-effects can be best studied when nearest-neighbor sequences are also accounted for. To investigate all of the 136 tetranucleotide fragments (ABCD), the molecular dynamics study was made of 39 double-stranded B-DNA oligomers, each containing 18 base pairs. The sequence of each oligomer was constructed in the same way: 5'-gc-CD-ABCD-ABCD-ABCD-gc-3', where upper case letters indicate sequences that vary and lower case letters indicate fixed sequences [19]. All simulation times were between 50 - 100 ns using explicit solvent and physiological salt concentration.

The results show that nearest-neighbor effects (the flanking base sequences) on base pair steps are very significant, which means that dinucleotide models are insufficient for predicting sequencedependent behavior. Some helical parameters do not necessarily have normal distribution [19]. This is the case for the minor groove width at the level of trinucleotide fragments. As shown in Figure 14, the groove width for A-T base pair has two possible states. For flanking sequence AAA the minor groove width is preferably narrow with typical values around 4Å. On the other hand wide minor groove width of 8Å is preferred by flanking sequence CAG. Some other flanking sequences, like GAG, exhibit a dynamic equilibrium between the two states.

Figure 14: Minor groove width distributions for A-T base pairs with different flanking sequences: AAA (blue), CAG (green) and GAG (red) [19].

At the level of tetranucleotide fragments the study showed that the extent of nearest-neighbor influence on helical parameters is defferent for different parameters [19]. Figure 15 illustrates how average values for some base-pair parameters are dependent on the flanking sequences. The examples are for one purine-purine (GG), one purine-pyrmidine (GT) and one pyrmidine-purine (CG) step. Tilt and roll (not shown) are hardly affected by nearest-neighbor sequence. Tilt is always small and roll, although variable, is largely determined by the base pair step itself. The nearestneighbor effects are more significant for shift and slide, where deviations are up to 60%. Changes occur mainly for purine-purine steps (GG, GA, AG and AA), but aside form this observation there are no apparent patterns. In contrast, rise and twist are also strongly affected with changes of up to 0.7Å and 18°, but, in this case, primarily for pyrmidine-purine steps (most notable are TG and CG steps).

Figure 15: Average values of inter-BP parameters for the unique base pair steps as a function of the flanking sequences. The three groups of four plots refer to the helical parameters of GG, GT and CG step. In each plot, the two-letter code along the abscissa indicates the 5' and 3' flanking bases and the horizontal line indicates the sequence-averaged value of the parameter. Translational parameters are given in angstroms and rotational parameters in degrees [19].

Standard deviations of twist for base-pair step CG are shown on Figure 16 for three distinct environments. Low twist is favored in CCGA and high twist in ACGT flanking sequences. On the other hand in ACGA environment the twist function is bimodal.

Figure 16: Distribution of CG twist (degrees) as a function of the flanking sequences: CCGA (blue), ACGT (green) and ACGA (red) [19].

7 CONCLUSION

With the rapid development of computer technology, allowing larger systems to be studied on a longer time scale and enabling also more accurate and detailed description of molecular interactions, the importance of molecular simulations in this area will grow even further. Furthermore, there are continuous advances in simulation methods. One of the promising one is the so-called adaptive resolution method where the atomistic and coarse-grained region are coupled within one simulation. Atomistic regime in the region of interest, let's say for the DNA molecule, gives us very detailed information, while the coarse-grained regime in the regions far away will significantly accelerate the computation.

Researching DNA is not attractive only from the theoretical aspect, but also form technological. Some scientists claim that DNA could be used for practical applications, such as to produce electronic devices like nanowires, or as parallel computers to solve very difficult combinatorial optimization problems [11].

References

- T. Schlick, Molecular Modelind and Simulation: An Interdisciplinary Guide, Pearson Education Limited (2001).
- [2] A. R. Leach, Molecular Modeling; Principles and Applications, Springer, New York (2010).
- [3] http://academic.brooklyn.cuny.edu/biology/bio4fv/page/molecular%20biology/dnature.html.
- [4] EMBO Workshop, Definitions and nomenclature of nucleic acid structure parameters, EMBO Journal 8, 1-4 (1989).
- [5] R. Podgornik, *Physics of DNA*, unpublished.
- [6] M. P. Allen, Introduction to Molecular Dynamics Simulation, Computational Soft Matter: From Synthetic Polymers to Proteins, NIC Series 23, 1-28 (2004).
- [7] W. M. Gelbart, R. F. Bruinsma, P. A. Pincus N. and V. A. Parsegian, DNA-Inspired Electrostatics, Physics Today 38-44 (2000).
- [8] A. P. Lyubartsev, Molecular Simulations of DNA Counterion Distributions, Encyclopedia of Nanoscience and Nonotechnology, Marcel Dekker, Inc., 120009094 (2004).
- R. C. DeMille and V. Molinero, Coarse-Grained Ions without Charges: Reproducing the Solvation Structure of NaCl in Water Using Short-ranged Potentials, J. Phys. Chem. 131, 034107 (2009).
- [10] A. P. Lyubartsev and L. Nordenskiöld, Monte Carlo Simulation Study of Ion Distribution and Osmotic Pressure in Hexagonally Oriented DNA, J. Phys. Chem. 99, 10373-10382 (1995).
- [11] T. A. Knotts, N. Rathore, D. C. Schwartz and J. J. Pablo, A Coarse Grain Model For DNA, J. Phys. Chem. 126, 084901 (2007).
- [12] R. C. DeMille, T. E. Cheatham III and V. Molinero, A Coarse-Grained Model of DNA with Explicit Solvation by Water and Ions, J. Phys. Chem. B 115, 132-142 (2011).
- [13] L. Nordenskiöld, N. Korolev and A. P. Lyubartsev, DNA-DNA Interactions, DNA Interactions with Polymers and Surfactants, John Wiley & Sons, Inc., New Jersey (2008).
- [14] D. C. Rau and A. Parsegian, Direct Measurment of Temperature-dependent Solvent Forces Between DNA Double Helices, Biophysical Journal 61, 260-271 (1992).
- [15] N. Grønbech-Jensen, R. J. Mashl, R. F. Bruinsma and W. M. Gelbart, Counterions-Induced Attraction between Rigid Polyelectrolytes, Phys. Rev. Lett. 78, 2477 (1997).
- [16] E. Allahyarov, G. Gompper and H. Löwen, DNA condensation and redissolution: interaction between overcharged DNA molecules, J. Phys.: Condens. Matter 17, 1827-1840 (2005).
- [17] A. P. Lyubartsev and L. Nordenskiöld, Monte Carlo Simulation Study of DNA Polyelectrolyte Properties in the Presence of Multivalent Polyamine Ions, J. Phys. Chem. B 101, 4335-4342 (1997).

- [18] N. Korolev, A. P. Lyubartsev, L. Nordenskiõld and A. Laakosonen, Spermine: An "Invisible Component in the Crystals of B-DNA. A Grand Canonical Monte Carlo and MOlecular Dynamics Simulation Study, J. Mol. Biol. 308, 907-917 (2001).
- [19] R. Lavery et al., A Systematic Molecular Dynamics Study of Nearest-neighbor Effects on Base Pair and Base Pair Step Conformations and Fluctuations in B-DNA, Nucleic Acids Research 38, 299-313 (2010).