Modeling Nucleic Acids at the Residue-Level Resolution

Filip Leonarski and Joanna Trylska

Abstract. Coarse-grained models and force fields have become useful in the studies of the dynamics and physicochemical properties of nucleic acids. Reduced representations of DNA or RNA allow saving computational cost of a few orders of magnitude in comparison with full-atomistic simulations. In this chapter we describe a few coarse-grained models of nucleic acids in which one nucleotide is represented as either one, two, or three beads. We selected the examples of the models designed to investigate the internal dynamics and temperature-dependent denaturation of nucleic acids, as well as created to predict the tertiary structure of RNA or used for large ribonucleoprotein complexes. We describe how the purpose of the model affects the design of the potential energy function and the choice of the simulation method. We also address the limitations of these models.

1 Introduction

Genomes of many species, including human, have been already mapped [1, 2] and are publicly available [3]. Their analyses give critical information on the cell components. However, in numerous cases, looking solely at the nucleotide sequence is not enough to explain how the processes in the cell are controlled. This happens because these sequences give rise to three-dimensional molecules, immersed in the environment of the cell, which undergo thermal fluctuations and "precisely" interact. Therefore, the knowledge of the sequence, even though crucial, is only the first step to analyze the spatial and temporal pattern of biomolecular interactions. To understand these interactions one needs to capture both the structural properties and

Filip Leonarski
Centre of New Technologies and Faculty of Chemistry,
University of Warsaw, Warsaw, Poland
e-mail: F.Leonarski@cent.uw.edu.pl

Joanna Trylska
Centre of New Technologies, University of Warsaw, Warsaw, Poland
e-mail: joanna@cent.uw.edu.pl

A. Liwo (Ed.): *Comput. Methods to Study the Struct. & Dyn. of Biomolecules*, SSBN 1, pp. 109–149.
DOI: 10.1007/978-3-642-28554-7_5

time-dependent dynamics of single molecules and macromolecular complexes. Below, we give a few examples where the dynamics is indispensable for biological function.

Cells use multiple strategies to pack and protect long strands of deoxyribonucleic acid (DNA) to provide for both DNA compaction and DNA accessibility for transcription, replication, and repair. For example, in bacteria DNA is supercoiled, a torsional stress is applied to a circular DNA duplex (plasmid) [4, 5]. The changes in supercoiling result in a bacterial response to hostile conditions such as starvation or thermal shock [6, 7]. The unwinding of supercoiled DNA is also the first step of transcription and replication [8]. In eukaryotes, proteins are used to help pack DNA in the nucleus to the form of chromatin. The simplest building block is the nucleosome, which is composed of histone proteins that are wrapped around by about 140 base pair long DNA duplex [9]. Multiple successive nucleosomes are separated by DNA linkers and resemble "beads on a string" under an electron microscope. Such organization allows the cell to control access to nucleosomal DNA, which is possible only when DNA unwraps from the histone core. Therefore, understanding the dynamics of this mechanism is crucial to control the gene expression or design how to put the genetic material into cells.

Another important aspect of the stability of DNA is related to the flexibility and dynamics of its double–helical structure. Topological stress, temperature or force–pulling might break bonds between the complementary bases and destroy the helix. DNA denaturation is easily tracked by UV–monitored changes of absorbance upon raising the temperature. This process depends on the sequence and length of DNA, and solution conditions such as ionic strength and pH [10]. Although in living cells a complete denaturation is not desirable, the local opening of a double–helix is important for gene regulation. A small "bubble" of denatured DNA forms in the areas where the transcription is initiated and/or regulated [8].

Though ribonucleic acid (RNA) differs from DNA by just one hydroxyl group in the sugar ring, this difference has important implications for the RNA architecture leading to a plethora of RNA structures with diverse roles. Messenger RNAs (mRNAs) serve as templates to transfer genetic information from DNA to ribosomes. The RNAs that do not carry genetic information form a large group of non–coding RNAs [11]. Transport RNAs supply the ribosome with amino acids. The ribosome itself contains ribosomal RNA which serves not only as a structural skeleton but also as a catalytic center. There is also a myriad of regulatory RNAs such as micro RNAs, small interfering RNAs, and small nucleolar RNAs.

Functional differences between DNA and RNA arise from the structural ones. DNA predominantly forms an ordered double–helical structure, with adenosine–thymine (A–T) and guanine–cytosine (G–C) complementary canonical base pairs. RNA is predominantly single–stranded with nucleotides bound by both complementary and non–complementary hydrogen bonds. Complementary ones, in the Watson–Crick sense, represent the secondary structure. They are formed first, in microsecond to millisecond time scales [12]. Bonds formed according to other schemes are responsible for the RNA 3D folds and the entire tertiary structure [13, 14]. The tertiary structure formation requires even seconds. The network of interactions in

RNA leads to double helical regions, intertwined with loops and junctions. Evolutionary conservation analyses show a strong link between the tertiary structure of RNA and its function, whereas the secondary structure and sequence are less conserved [15].

The functionality of RNAs is related to its flexibility and ability to change folds [16]. Some RNAs adapt multiple functional conformations in response to external conditions. The examples are riboswitches [17] which respond to ligand or metal ion concentrations and RNA thermometers [18] which respond to temperature shifts. These are mRNA fragments that typically include the Shine–Dalgarno sequence [19] responsible for binding of mRNA to the ribosome and initialization of translation. The Shine–Dalgarno sequence either forms a hairpin loop which is not exposed to interact with the ribosome or it switches the fold and the sequence becomes accessible to the ribosome. The accessibility of Shine–Dalgarno sequence depends on the environment and can be moderated by external conditions.

Full understanding of the above processes requires the knowledge of how the structure of nucleic acids changes and fluctuates in time and how this dynamics is related with function. The methods that gain information solely from sequence are of great value, e.g., thermodynamic nearest–neighbor model has been successful in predicting denaturation temperatures of various DNA or RNA duplexes [20, 21, 22]. Also, the secondary structure can be in most cases reliably predicted just based on the sequence [23]. However, the sequence-based methods fail for more complicated tasks such as predictions of RNA 3D structure [24] and more importantly dynamics. The dynamics, which is typically simulated based on the 3D model, helps in understanding the functional roles of various nucleic acid architectures.

Also, in comparison with the number of available sequences of functional nucleic acids, the experimentally-determined 3D structural data lag behind. As of June 2012, there have been 928 RNA structures deposited in the Protein Data Bank [25]. In the year 2011 there were only 72 new RNA structures resolved. When compared with proteins these numbers are 75708 deposited structures and 7547 resolved in 2011. Efficient ways to predict the RNA 3D structure will help filling the gap of low number of RNA structures in the crystallographic database. The dynamical data for RNA are even more sparse also because the dynamics is difficult to be monitored experimentally at atomic level and on fast time scales. So the modeling methods that add the fourth dimension — time-dependence are beneficial to understand the complexity of interactions in the cell.

To characterize the dynamical processes occurring in nucleic acid molecules multiple techniques have been used. The three main conformational sampling techniques, are molecular dynamics simulations, normal mode analysis, and Monte Carlo (MC) algorithms [26]. All typically require, as a starting point, a set of initial coordinates of the molecule describing its 3D structure. The MC algorithms are probabilistic methods that help to stochastically explore the conformational space of molecules. In an MC simulation (e.g., [27, 28]) small modifications of molecule's coordinates are randomly introduced and are either accepted or rejected based on the potential energy of the system. If the modification lowers the potential energy, it is always accepted. Otherwise, its acceptance is probabilistic, more likely to happen

if an absolute value of energy change is small. A wide number of possible conformations can be probed using this method. Conversely, if one is not interested in a wide search of conformational space but in low frequency dynamics of a known native state, normal mode analysis can be applied [29]. Normal mode analysis predicts the system's motions at equilibrium by decomposing them into independent vibrational modes. This method looks for vibrational normal modes with lowest frequencies which are usually connected with molecule's function. Molecular dynamics (MD) [30, 31] is a tool most often used to analyze the time-dependent dynamic behavior of biomolecules. By integrating Newton's equations of motion one can calculate the positions and velocities of atoms or residues at small subsequent time steps (the trajectory). However, to solve these equations one has to provide initial positions and velocities. The latter ones are usually assigned according to the Boltzmann–Maxwell distribution at a requested temperature.

All these methods require a mathematical formula with a set of parameters (force field, FF) to calculate the potential energy of the system. Well-known examples of such FFs are Amber [32, 33] or CHARMM [34, 35], which provide sets of parameters to simulate proteins, nucleic acids, lipids, and other molecules. They employ a full-atomistic representation of a molecule, i.e., consider each atom separately in integrating equations of motion. To provide a good description of the environment one has to include solvent effects. This can be achieved by explicitly adding water (or other solvent) molecules to a system. The state-of-the-art examples of MD simulations in explicit solvent include a millisecond simulation of a 58 amino-acid bovine pancreatic trypsin inhibitor protein, performed on a computer build exclusively and purposely for MD simulations by D. E. Shaw group [36, 37] and 13.3 μs MD simulation of folding of a 162 amino-acid human pin1 WW domain by K. Schulten group [38]. Traditional full-atomistic FFs have their limitations primarily because they were parameterized based on experiments and quantum-mechanical calculations for small molecules. There are also doubts about the quality of the microsecond scale simulations since the FF parameterization was not performed with such long time scales in mind. Unfortunately, the microsecond time scale is still too short to model global conformational changes in RNA, to fully grasp how the RNA tertiary structure is formed or to predict unliganded states of riboswitches.

Time saving by at least an order of magnitude can be achieved by performing a simulation with the solvent modeled implicitly. To do this one can modify the FF to include hydrophobic effects by adding a term involving the solvent accessible surface area [39]. One can also modify equations of motions to include random collisions with water, like in the Langevin-type dynamics [40]. But simplifying a system can go further than just removing the solvent degrees of freedom. One can reduce the system's representation to achieve the necessary reduction in complexity. In such simulations chemical groups or even whole residues can be represented as single interacting centers (beads). Then the gain in performance is two-fold. The more obvious one is the decrease of the number of interactions in the calculations of the potential energies or forces. Additionally, the most frequent vibrations are removed from the system, smoothing the potential energy surface, and allowing one to use a larger simulation time step. Therefore, such coarse-graining (CG)

procedure, should be appropriate to simulate the above introduced nucleic acid dynamical problems that occur on nanoseconds to seconds.

In this chapter we present the CG FFs for nucleic acids that use between one and three beads per nucleotide. We believe that such models give a reasonable balance between the quality of the results and time efficiency of the calculations. However, there are models that use a higher number of beads and represent the structural details of bases (e.g., [41, 42, 43, 44, 45, 46, 47, 48]). On the other spectrum there are coarser models in which the building blocks are formed of helices and single-stranded loops (not single nucleotides) [49, 50, 51, 52, 53]. Here, we describe only the models that use spherical beads but some authors implement interaction centers as ellipsoids [54] or disks [55]. We will also not cover the two models which are historically important: a one-bead model published in 1970s by W. K. Olson [56, 57, 58] which was the first attempt of coarse-grain DNA modeling and a three-bead per nucleotide model by Y. N. Vorobjev [59] from 1990, since the latter model was not used in actual simulations, despite its strong theoretical background. Also, the models that we review here belong to the class of the off-lattice models for which the bead coordinates in the simulations are not limited to a certain set of positions such as nodes of a cubic grid. Our selection of the models is arbitrary and far from complete because our aim was to give informative examples of how the CG models for nucleic acids are constructed and to which biological problems they can be applied.

2 Coarse-Grained Force Field Parameterization

Coarse-graining is not a problem-free procedure. One of the challenges of the CG models is their parameterization. For full–atomistic models there are well–established protocols where FF parameters are determined and benchmarked based on quantum chemistry models and experimental measurements of thermodynamic parameters for small molecules. CG models are hard to fit directly to the quantum mechanical data, although there are examples such as the UNRES FF developed by Liwo et al. [60, 61] (and described in this book) where tedious derivations led to a usable CG potential.

Most CG potentials presented in this chapter can be classified as statistical or knowledge–based. The parameters of such FFs were found based on the average properties derived from large sets of reference data characterizing the molecules of interest. In some cases, the data sets include all nucleic acids of a certain class found in one of the crystallographic databases [25, 62]. In other cases, the data sets are gathered from full–atomistic simulations. However, in both cases the parameterization procedure is the same. Using the Boltzmann inversion procedure one can infer the potential energy from distributions of certain observables (e.g., distances, angles, dihedrals) acquired from one of the mentioned sources [63]. A distribution $d(r)$ of an observable r is linked with the potential energy using the equation

$$V(r) = -k_B T \ln \frac{d(r)}{d_0(r)} , \tag{1}$$

where $d_0(r)$ is a reference distribution of such an observable, k_B is the Boltzmann constant, and T temperature. To use this method one has to assume that $V(r)$ is not correlated with other observables, which, if not satisfied, can lead to errors in the potential energy approximation. However, there are also other problems associated with the Boltzmann inversion approach. The structural data set might be biased (e.g., the PDB database contains rather short RNA molecules and long RNAs are underrepresented) or affected by uncertainties in the structure determination due to low resolution of electron density maps. Also, the structures of biomolecules from X–ray crystallography are derived at temperatures much lower than 310K. Moreover, crystallization of biomolecules typically occurs under unphysiologic high-salt conditions and induces crystal packing forces. Last, but not least, finding the proper reference distribution, $d_0(r)$, is difficult because it should take into account the specifics of nucleic acid structures (such as the linearity).

Therefore, some authors fix certain parameters to the values that are experimentally known, e.g., from the thermodynamic measurements. Next, this procedure is followed by a trial–and–error optimization of other parameters in order to correct for the drawbacks of the Boltzmann inversion method. Typically, one performs tests of a CG FF on a known system and systematically modifies the parameters until a reliable set meeting the assigned criteria is found. This last step can be performed in a systematic way using local or global optimization methods [64, 65, 66, 67].

3 Force Field Description

The basic criterion which we apply to divide the CG models into classes is the number of beads used to represent one nucleotide. As stated in the introduction we will cover only a small spectrum of possible representations — one to three beads per nucleotide. However, even in this bead range one observes differences between the design and applicability of the coarser– and finer–grained models.

There are also other measures to compare FFs apart from the number of interacting centers. These are the mapping (where the centers of beads are positioned), the definition of potential energy function, and the range of applicability (or transferability). Unfortunately, for the CG methods this applicability range is usually very narrow. To provide reliable and predictive results CG methods have to be fine–tuned for a particular process and/or group of molecules. Overall, CG FFs lack the general transferability of all-atom ones. For example, in the presented set of FFs there is not even one that can be, out of the box, applied to both RNA and DNA systems. Apart from the target molecule, we have selected the following main classes of problems that the FF can be applied to:

- Long timescale dynamics — a model provides reliable information about the time evolution of a molecular structure on at least nanosecond scale.
- Tertiary structure prediction — a model finds a 3D structure or a set of 3D structures that are closest to the native state. The focus is only on the final structure, not on the way it is achieved (in contrast to the folding simulations).

- Temperature denaturation — a model correctly predicts the effects of the temperature increase on nucleic acid stability.
- Supercoiling — a model predicts the effects associated with supercoiling (e.g., unwinding mechanics).
- Large molecule mechanics — a model is designed to simulate the dynamics of large molecular complexes ($>$ 1000 residues) such as the ribosome or nucleosome.
- Interaction with non–nucleic acid molecules — a model is able to predict interactions with ligands, proteins or nanomaterials (ions and solvent are not included in this category).

The FF applicability results from its implementation details such as the definition of the potential energy function with respect to the chosen degrees of freedom and connectivity. For the residue-resolution CG FFs the potential energy function, V_{total}, is usually expressed in the following, general way:

$$V_{total} = V_{intrastrand} + V_{interstrand} + V_{nb} \,. \tag{2}$$

The *intrastrand* term covers the interactions of beads connected by covalent bonds which extend up to the third neighbor. This term is composed of a pseudo–bond (V_{bond}), pseudo–angle (V_{angle}), and pseudo–dihedral ($V_{dihedral}$) parts (see Fig. 1):

$$V_{intrastrand} = V_{bond} + V_{angle} + V_{dihedral} \,. \tag{3}$$

Typically, these bonds are not allowed to break in a simulation, so they are represented with harmonic potentials (see Fig. 1a,b,c):

$$V_{bond}(r) = k_r(r - r_0)^2 \,, \tag{4}$$

$$V_{angle}(\theta) = k_\theta(\theta - \theta_0)^2 \,, \tag{5}$$

$$V_{dihedral}(\phi) = k_\phi(\phi - \phi_0)^2 \,, \tag{6}$$

where k_r[1], k_θ and k_ϕ are the force constants, r_0 the equilibrium distance, and ϕ_0 and θ_0 are the equilibrium angles. The drawback of the above $V_{dihedral}$ is that it is not periodic so to account for full rotation of the pseudo–dihedral angle, a formula with a cosine is used (see also Fig. 1c):

$$V_{dihedral}(\phi) = k_\phi[1 - \cos(\phi - \phi_0)] \,, \tag{7}$$

with the same definition of k_ϕ and ϕ_0. Beads positioned in the same strand can form complementary bonds which is especially important for RNA that is usually

[1] In this chapter we ignore the $\frac{1}{2}$ factor because the harmonic potentials in CG FFs are presented differently (either with or without the $\frac{1}{2}$ factor). Including this factor affects only the numerical value of a force constant but does not change its general form.

composed of only one folded strand. However, as these are usually residues separated by more than three bases, for the purpose of CG FFs such bonds are not considered to be "intrastrand" and accounted for in the interstrand part.

The *interstrand* term describes the interaction of complementary strands. This term models hydrogen bonds which in nature can be broken by raising the temperature or adding denaturating agents or enzymes. Breakable bonds are usually implemented using the Lennard–Jones potential (see Fig. 2a):

$$V_{LJ}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right], \tag{8}$$

or in an alternative form ($r_0 = 2^{1/6}\sigma$):

$$V_{LJ}(r) = \varepsilon \left[\left(\frac{r_0}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^{6} \right], \tag{9}$$

or the Morse (see Fig. 2b) potential:

$$V_{Morse}(r) = V_0(\exp[-\alpha(r - r_0) - 1]) - 1)^2 - V_0 \,. \tag{10}$$

Equations (8) and (9) are two forms of the same equation. ε describes the depth of the potential energy well. σ is the distance where the potential energy is equal to zero and r_{eq} is the distance where the potential energy has a minimum. For the Morse potential of Eq. 10, V_0 is also the depth of the energy well and α describes the width of the potential well. The Lennard–Jones potential might be modified (for example softened) by changing the powers in the equation. However, not all FF models permit such actions because this requires a more complex potential energy formulation. It is not always necessary to allow for the interstrand bond breaking because a particular CG model may be designed only for non–denaturating conditions. In such case a simple (4) harmonic potential may suffice. The CG models also differ in the way the interstrand bond network is set. Simpler models have a predefined network which is based on the secondary structure prediction and the pairing is not altered during a simulation, so even after denaturation the molecule will always return to the same conformational setting as in native conditions. This is beneficial for RNA structure prediction, when we are interested in the folds that correspond only to one particular secondary structure. In the case of more elaborate CG FF models interstrand bonds can be formed dynamically when the two complementary bases are close and their topology permits bonding.

The last category of terms are the nonbonded ones *nb*. They account for the interactions of residues that are not connected explicitly by intrastrand and interstrand terms. Their basic function is to introduce a short–range repulsion to avoid overlapping of non–interacting beads, however, they also account for long-range electrostatic interactions and solvent or other environmental conditions. The implementation of these terms varies among FFs depending on their applications. Some FFs use Lennard–Jones or Morse terms as in (8) or (10) that describe both the attraction at short distances and repulsion at long distances. However, for highly

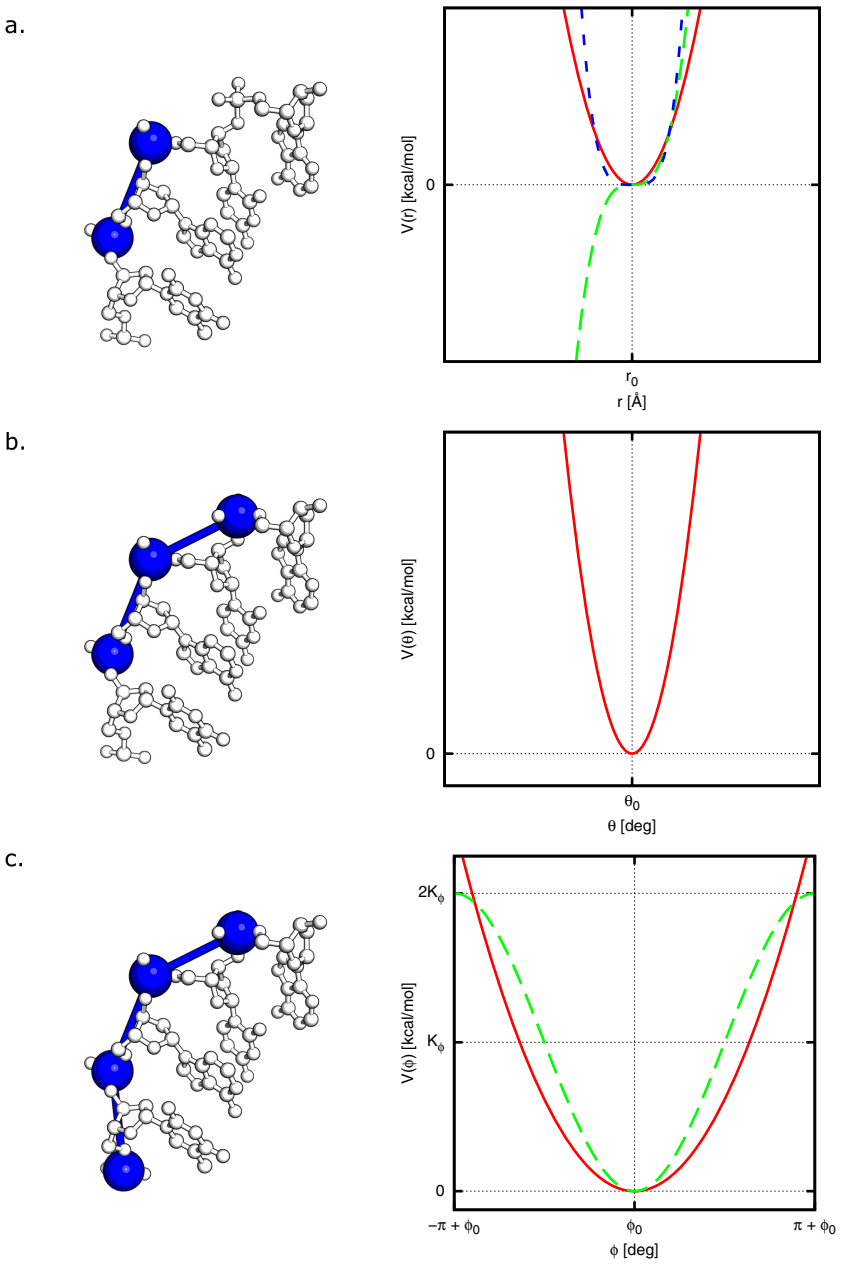

Fig. 1 Intrastrand potentials used in the presented CG FFs. **a.** The pseudo–bond harmonic (*solid line*,see (4)), cubic (*long–dashed line*) and quartic (*short–dashed line*, see (33)) potential **b.** The pseudo–angle potential (see (5)) **c.** The pseudo–dihedral potential implemented using a cosine function (*long-dashed line*, see (6)) or harmonic potential (*solid line*, see (7)).

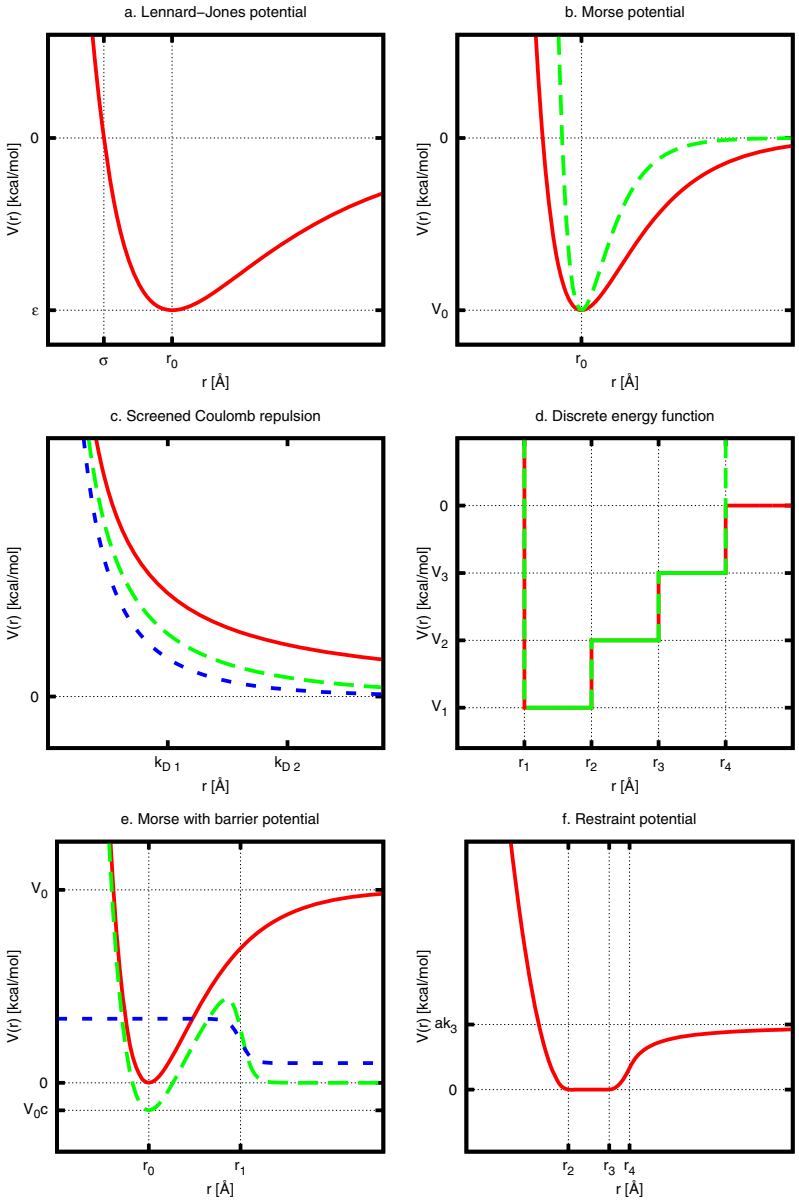

Fig. 2 Interstrand and nonbonded potentials used in the presented CG FFs. **a.** Lennard–Jones potential **b.** Morse potential with $\alpha = 1.0$ (*solid line*) and $\alpha = 2.0$ (*long–dashed line*) **c.** Coulomb potential without screening $k_D = \infty$ (*solid line*), Coulomb potential with two example Debye lengths $k_{D1} < k_{D2}$ (*short– and long–dashed line*, respectively) **d.** Discrete potential taken from the model of Ding et al. [69, 70] **e.** Morse potential with a barrier used in Trovato et al. [71]: Morse potential (*solid line*), switch function (*short–dashed line*), final potential (*long–dashed line*) **f.** Restraint potential from the model of Malhotra et al. [72, 73, 74].

charged molecules, such as nucleic acids, one could also use the Coulomb electrostatic potential to describe the repulsive—only potential, with or without shielding (see Fig. 2c):

$$V_{Coulomb}(r) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_w r}\,, \tag{11}$$

$$V_{ShCoulomb}(r) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_w r}\exp(-r/k_D)\,, \tag{12}$$

where q_i and q_j are the charges of interacting beads, ε_0 is the vacuum and ε_w the solvent permittivity. The Debye length,

$$k_D = \left(\frac{\varepsilon_0\varepsilon_w k_B T}{2N_A e^2 I}\right)^{0.5}\,, \tag{13}$$

depends on the temperature T and ionic strength I of the solution. k_B is the Boltzmann constant, N_A is the Avogadro number, and e is the electron charge [68].

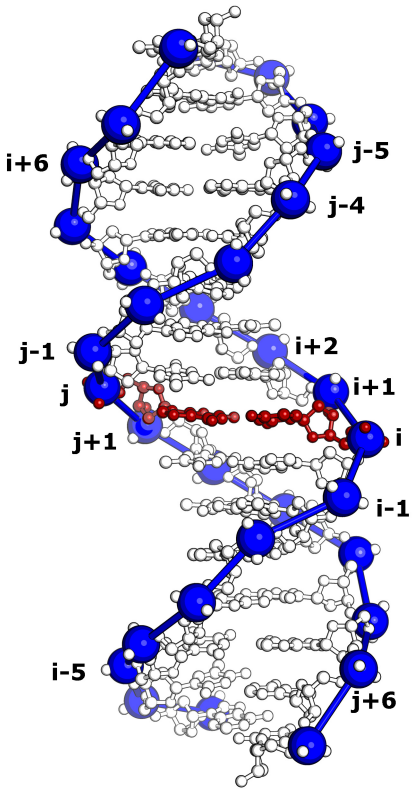

Fig. 3 DNA helix showing the nucleotide numbering according to the *i*:*i*+*n* and *i*:*j*+*n* convention, with a single nucleotide pair (*darker*) in the middle as a reference. This helix is shown in a one–bead representation, with interaction centers placed on phosphorus atoms as in FF by Trovato and Tozzini [71] and Trylska et al. [75, 76].

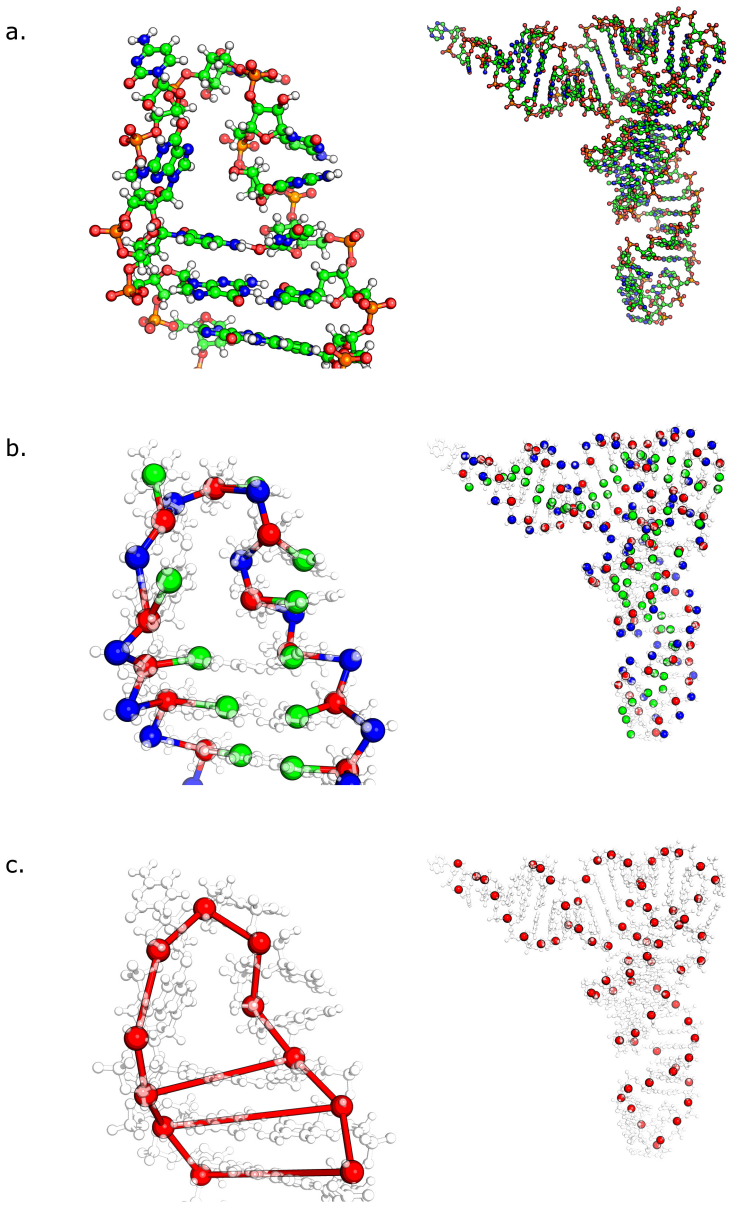

Fig. 4 Left: RNA hairpin loop (PDB:1ATO [77]); **Right:** yeast phenylalanine tRNA (PDB:6TNA [78]): **a.** full–atomistic representation **b.** three–bead per nucleotide representation as in the work of Ding et al. [69]. **c.** one–bead per nucleotide as in the work of Jonikas et al. [79]. For the RNA hairpin loop (left) we show the bead placement with non–breakable bonds and for tRNA (right) we show only the bead placement.

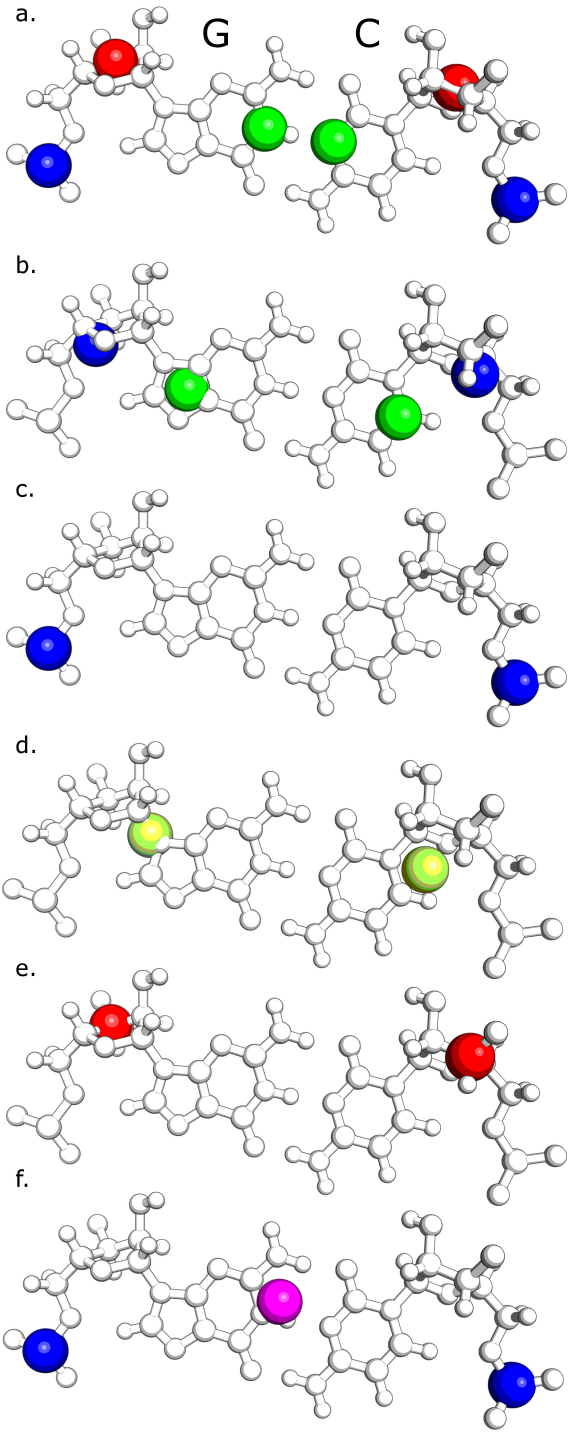

Fig. 5 Guanine – cytosine nucleotide pair represented in different CG representations: **a.** three–bead model as in Knotts et al. [68], a similar model is described in the work of Ding et al. [69], however the base atom is placed in the center of the 6–member nucleotide ring. **b.** a two–bead model with pseudo–atoms placed on the backbone and base as in Drukker et al. [80] **c.** one bead centered on the phosphorus atom as in Trovato et al. [71] and by Trylska et al. [75, 76] **d.** one bead placed in the nucleotide geometric center as in Savalyev et al. [67] **e.** one bead centered on the C3' atom as in Jonikas et al. [79] **f.** one bead placed on the phosphorus atom and a special "dummy" bead in the middle of a complementary pair as Malhotra et al. [72, 73, 74]

Table 1 Comparison of features of CG FFs presented in this chapter

	Knotts et al. [68, 81, 82, 83]	Ding et al. [69]	Hyeon et al. [84]	Ouldridge et al. [85, 86, 87]	Drukker et al. [80]	Trovato et al. [71]	Savalyev et al. [67, 88]	Jonikas et el. [79]	Trylska et al. [75, 76]	Malhotra et al. [72, 73, 74]
Number of beads / nt	3	3	3	3	2	1	1	1	1	1.5
Nucleic acid										
DNA	✓			✓	✓	✓	✓		✓	
RNA		✓	✓					✓	✓	✓
Applicability										
Long timescale dynamics	✓	✓	✓	✓	✓	✓	✓		✓	✓
Tertiary structure prediction		✓						✓		
Temperature-dependent denaturation	✓		✓	✓	✓	✓				
Force–pulling denaturation			✓							
Supercoiling						✓	✓			
Large macromolecule mechanics									✓	✓
Interaction with non–nucleic acids									✓	✓
Potential energy formulation properties										
Bias toward reference positions			✓						✓	✓
Bias toward reference secondary structure		✓	✓			✓	✓	✓	✓	✓
Base specificity	✓	✓	✓	✓	✓					
Breakable hydrogen bonds	✓	✓	✓	✓	✓	✓				
Explicit ions	✓						✓			

For intrastrand and interstrand interactions we introduce the following notation shown in Fig. 3: *i:i+n* denotes the interaction between a nucleotide and its *n*-th successor on a single strand, *i:j+n* or *i:j-n* denotes the interaction between a nucleotide and its *n*-th successor (or predecessor) of its complementary strand.

A graphical representation of CG FFs described in this chapter can be found in Figs. 4 and 5. In Tab. 1 we compare the features of the described models. In the following sections we present the models in the descending order of complexity — from three to one bead per nucleotide.

4 Three-Bead DNA Model for Dynamics and Melting

The first example that we describe of a three–bead per nucleotide model is the one of Knotts et al. [68] designed for DNA. In this model the beads that mimic the sugar and phosphate are placed at the centers of mass of these groups. The adenine and guanine base beads are placed in the position of their N1 atoms and the thymine and cytosine beads in the position of their N3 atoms (see Fig. 5a). The authors argue that representing the DNA backbone with two beads is necessary to properly model the deformation of grooves which are important for protein—DNA interactions. The choice of a three–bead representation also helps in later transformation from a CG representation to a full–atomistic one. The intrastrand part of the potential energy function contains one additional term, V_{stack}, in comparison with (3):

$$V_{intrastrand} = V_{bond} + V_{angle} + V_{dihedral} + V_{stack} \,. \tag{14}$$

The pseudo–bond V_{bond} and pseudo–angle V_{angle} potentials are implemented using harmonic potentials (see (4) and (5) and Fig. 1a and Fig. 1b). The pseudo–dihedral potential $V_{dihedral}$ is implemented using a cosine potential (see (7) and Fig. 1c). The V_{stack} term is modeled with the Lennard–Jones potential (see (8) and Fig. 2b).

The first three terms in (14) are standard but V_{stack} is an additional Go–type potential introduced to account for the stacking interactions [89]. This interaction is modeled only between the base beads that belong to one strand and the reference ("native") structure is positioned within a 9 Å cut–off distance. Therefore, this potential accounts for both the *i:i+1* and *i:i+2* interaction.

In the interstrand term the complementary base pairs are connected using the Lennard-Jones-like potential (see Fig. 2a), but with the 12–10 powers instead of 12–6 as in (8):

$$V_{interstrand}(r_{ij}) = 4\varepsilon_{bp_{ij}} \left[5 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{10} \right] , \tag{15}$$

where the summation is over all G–C and A–T base pairs that are not already considered in V_{stack}.

The nonbonded potential in the original paper [68] is composed of an excluded volume term V_{ex}, implemented using the Lennard–Jones potential (see (8) and Fig. 2a) and a shielded electrostatic term $V_{ShCoulomb}$ (see (12) and Fig. 2c):

$$V_{nb} = V_{ex} + V_{ShCoulomb} \,, \tag{16}$$

where the V_{ex} term is only calculated when the r_{ij} distance between beads is smaller than a predefined cut–off. The $V_{ShCoulomb}$ defines the electrostatic repulsion of only phosphorus atoms (with the charges $q_i = q_j = -1$).

This model was parameterized in an iterative way. The first guess of parameters was taken from the geometry of an ideal B–DNA helix. Second, a 14 base–pair DNA duplex was simulated with the CG model using replica-exchange MD [90]. Eight replicas (or system copies) were simulated in parallel and assigned temperatures in the range 260–400 K. Temperatures were swapped between two replicas with a probability related to their potential energy difference. Each replica was equilibrated and 10 ns production runs were performed. The advantage of replica-exchange MD over constant-temperature MD was that it allowed the authors to determine the melting curves of the duplex and provided distance distributions in eight different temperatures. Also, the effect of parameters on the potential of mean force with varying temperature was analyzed using a weighted histogram analysis method [91] and the parameters were improved for the next iteration step.

Next, to validate the model, the obtained FF parameter set was evaluated by performing CG replica–exchange MD simulations and comparing them with the DNA thermal denaturation experiments. In the simulation the melting and the formation of the denaturation bubble were observed in accord with the reference data for varying salt concentrations. Knotts et al. [68] show that with their FF they were able to predict the melting temperatures of three DNA duplexes with an error lower than 5%. To validate the mechanical properties of the model, a CG traditional MD was performed at 300K. The persistence length for four different fragments of λ-phage plasmids (one of them was 1489 base pairs and 0.5 μm long) was calculated. Their model overestimated the persistence length by 2.3 but the authors claim that this is much less than in other CG models. Based on their parameterization Knotts et al. suggest that the dihedral force constant (k_ϕ), potential energy well depths for base–pairing ($\varepsilon_{bp_{ij}}$), stacking, and excluded volume (E_{ex}), are the most important parameters to tune.

The presented model was further improved. Sambriski et al. [92] added entropic effects to the potential energy to allow for rehybridization of the DNA strands, as the original model of Knotts et al. [68] was unable to model strands' renaturation. DeMille et al. [93] added explicit solvation with water as well as monovalent ions. This modification provides a good cylindrical distribution of ions around DNA but it over-estimates the DNA melting temperatures. Next, Freeman et al. [82] added to the model terms for the interactions of DNA with both mono– and di–valent ions.

This model is one of the most comprehensive CG FFs from the ones presented in this chapter. It can be used to estimate both DNA melting curves and DNA mechanical properties. The subsequent modifications of this model add better treatment of solvation and electrostatics. Nevertheless, there is still room for improvement, especially to correct for high errors of the calculated persistence lengths.

5 RNA Folding with a Three–Bead Model

The model by Ding et al. [69] was designed to predict the tertiary structure of RNA but may be also used to study the mechanism of RNA folding. This model is based on discrete MD previously successfully applied to protein folding [70, 94]. In this method, the interaction between beads is described using pairwise, discontinuous functions (see Fig. 2d):

$$V_{bond}(r) = \begin{cases} \infty & r < r_1 \\ V_1 & r_1 < r < r_2 \\ V_2 & r_2 < r < r_3 \\ \dots & \\ \infty & r > r_{max} \end{cases}, \tag{17}$$

$$V_{nb}(r) = \begin{cases} \infty & r < r_1 \\ V_1 & r_1 < r < r_2 \\ V_2 & r_2 < r < r_3 \\ \dots & \\ 0 & r > r_{max} \end{cases}. \tag{18}$$

Multiple–step distances $r_1, r_2, r_3, \ldots$ between beads are defined. If the distance between two beads is between r_1 and r_2 their pair potential interaction energy has a value of V_1, if this distance is between r_2 and r_3 the potential is assigned a different value – V_2, etc. If the distance is smaller than a minimal distance, then an infinite value of the potential is assigned to avoid overlapping. However, if the distance is larger than some maximal value, there are two possibilities; the potential energy is equal to 0 (if the interaction is considered "breakable") or infinity (if the interaction is considered for "unbreakable"). The functions described by (17) and (18) could not be used in traditional MD because of their discontinuity so Ding et al. [69, 70] have chosen a different approach. In principle, the bead velocities are constant during the dynamics and are changed only by colliding with other interacting centers. If bead kinetic energy is larger than the difference between the two energy steps $V_i - V_{i-1}$ and the distance is smaller than r_i, a collision can occur and velocities are updated. However, if the kinetic energy of beads is lower than a barrier, a hard reflection occurs without any change in the potential energy. The advantage of using this discrete MD method is its higher efficiency in comparison to standard MD. In the latter each MD step requires recalculating the forces acting on all atoms in the system and then solving the equations of motion. In the discrete method, in the case of no collisions, one needs to update only the positions of the beads, not velocities.

In this model single beads are assigned to a phosphate group (P), sugar (S), and base (B) (see Fig. 4b). As in the model of Knotts et al. [68] for DNA, the sugar and

phosphate beads are placed in the centers of masses of these groups and the base bead is placed in the center of a six–membered ring. The intrastrand interactions contain only the V_{bond} distance–dependent term and are a combination of unbreakable bonds between the P, S and B beads. Since there are no explicit pseudo–angle and pseudo–dihedral terms as in (3), additional bonds between the beads of two neighboring nucleotides are added (e.g., a bond between the S bead of an $(i-1)$–th nucleotide and a cytosine B bead of an i–th nucleotide). The stacking interaction between the bases is also implemented as a breakable bond (18) and designed in a way to provide a correct angle between the three bases in one line.

The interstrand terms are composed of breakable bonds between complementary nucleotides (also including the wobble pair G-U). A complementary pair is represented as three bonds: base–base and two sugar–base bonds. Such bonds are assigned only if a correct (in the Watson-Crick sense) distance and orientation between the sugar and base beads of both nucleotides are achieved. But in the case of loops the reduction of the degrees of freedom underestimates the entropy so loop forming may be modeled in an unphysical fashion. To account for better representation of loops, first, loop forming free energies are calculated according to the nearest–neighbor model [95] and for loops the interstrand bond is formed only with a probability based on this free energy value.

The nonbonded interactions are implemented as follows. The phosphate-placed beads repel each other by a discretized screened Coulomb potential (see Fig. 2c for the Coulomb potential and Fig. 2d for the general discrete potential). The base-placed beads are connected with an attractive force due to the hydrophobic nature of the nucleotides. The attraction between bases may result in overpacking of the bases, so there is an additional term which penalizes the bases with too many contacts in the defined cut–off region.

The model of Ding et al. [69] was parameterized based on the thermodynamic data from the nearest–neighbor model by Mathews et al. [95] and on distributions calculated from known 3D RNA structures. It was next evaluated on 153 known RNA structures of the lengths between 10 and 100 nucleotides. Their sequences were used to create linear RNA molecules, which were simulated with the discrete MD method [70] and their folding was analyzed. The so-called Q-values, defined as a fraction of native base pairs present in a given RNA conformation, were assessed. The average Q-value for all the tested structures was 94%, which is 3% higher than Mfold [96], a secondary structure prediction software (especially in the case of pseudo–knots). 84% of RNA structures had a root mean square deviation from the final reference structure lower than 4 Å, which is a good score. RNA folding with this potential can be performed using the iFoldRNA web server [97].

6 RNA Thermal Unfolding and Stretching with a Three–Bead Model

Another example of a three–bead per nucleotide FF is given by Hyeon et al. [84] and is an extension of a model that was previously designed for protein folding [98].

This RNA FF was created to model mechanical unfolding of a particular 22-nucleotide long RNA hairpin (P5GA hairpin) with a known NMR structure [99]. This hairpin is structurally similar to another P5ab hairpin of group I intron in the *Tetrahymena thermophila* ribozyme for which the force unfolding studies were performed using optical tweezers [100, 101]. Hyeon et al. [84] compare their simulation results of the P5GA hairpin to the ones from the above-mentioned experiments. Their CG model assigns beads to phosphate, sugar and base groups and places the beads in the geometrical centers of these groups. To create the topology the authors used the concept of the Go model [102] in which the interactions present in the native structure are attractive and all the others are repulsive.

The intrastrand potential, similar to the one used by Hyeon et al. [98] for proteins, is composed of three potential terms, like in (4), where the V_{bond} and V_{angle} terms use a harmonic function (see (4) and Fig. 1a and (5) and Fig. 1b) and $V_{dihedral}$ is implemented using the cosine potential (see (7) and Fig. 1c).

The interstrand potential is composed of a stacking term:

$$V_{stack} = \Delta G_i(T) F_{or} \,, \tag{19}$$

where $\Delta G_i(T)$ are the Turner's parameters of the nearest–neighbor model [95]. F_{or} is an orientation term, including both *i:j* and *i+1:j-1* distances and sugar and base bead angles involving *i*,*i+1*,*j*,*j-1*-th nucleotides (according to the *i:j* notation shown in Fig. 3).

The nonbonded term is described using the Lennard–Jones potential (see (8) and Fig. 2a), with separate formulas V_{nb}^{native} for the interaction of beads forming the native contacts (closer than 7 Å in the reference structure) and $V_{nb}^{non-native}$ for the interactions of non–native beads, and Debye-Huckel potential V_{PP} for the repulsion of phosphorus beads (see (12) and Fig. 2c):

$$V_{nb} = V_{nb}^{native} + V_{nb}^{non-native} + V_{PP} \,. \tag{20}$$

The FF was first tested by performing MD simulations of the unfolded P5GA, hairpin structure, without force steering, to see if the structure converges toward the NMR resolved one. By slow cooling, simulated annealing, and steepest–descent minimizations, the RNA hairpin converged to the experimentally folded one with the root mean square deviation of 0.1 Å. Next, the dynamics of stretching of the RNA P5GA loop was studied to calculate the phase diagrams of denaturation arising from external force and temperature. Finally, in the simulation, the hairpin was force pulled and later refolded from an extended conformation using a force quench. These MD simulations gave insight into the mechanism of force unfolding and refolding of the P5GA loop.

This model was further used to investigate the folding of RNA pseudo–knots. In the work by Cho et al. [103] the simulations of folding of three pseudo–knots (MMTV and SRV-1 from viral genomes and hTR from human telomerase) were performed and the folding mechanisms were consistent with experimental data. However, the authors emphasized that even though these pseudo–knots are structurally similar, their folding occurred through different scenarios. In the work by

Biyun et al. [104] further analysis of the hTR pseudo–knot folding was performed — the effects of the ion concentration jumps and temperature decrease on folding were investigated giving a better understanding of the transient states and folding pathways.

7 DNA Nanodevices with a Three Collinear Bead Model

The purpose of this three-bead model designed by Ouldridge and coworkers was to simulate the dynamics of DNA nanodevices [85, 86, 87]. The interactions in such DNA nanostructures are based on selective binding of complementary nucleotide pairs. DNA strands can be designed to form two-dimensional lattices [105], polyhedra [106, 107], or other regular structures [108]. There are also DNA structures in which the complementary hydrogen bonds are dynamically formed and broken. Overall, one can design a set of interacting DNA strands with a particular purpose in mind. A cycle based on single– to double–stranded DNA and reverse transitions may be used to create DNA tweezers [109] or DNA walkers that perform a directional movement on a DNA track [110, 111, 112]. To simulate such devices a CG model needs to correctly predict the complementary bond breaking and forming events. To satisfy this crucial requirement Ouldridge and coworkers have chosen a top–down methodology. In contrast to other models presented in this chapter, which are designed by mapping the full–atomistic structure on a CG set of positions, this model was designed in order to fit with the DNA hybridization and thermodynamic data. It might appear strange that the model ignores such basic measures as different sizes of DNA grooves. Its efficiency, however, is measured by the correspondence with hybridization enthalpies and entropies. And as long as there is an agreement between thermodynamic predictions and the 3D model, the model is considered acceptable for a particular task it was designed for.

In this FF a nucleotide is modeled as three collinear beads (see Fig. 6). A single bead mimics the position of the backbone and two beads represent a base — the first one is responsible for stacking and the second one is responsible for hydrogen–bonding and excluded volume interactions.[2] The distances between the backbone bead and base sites and between two consecutive backbone sites were chosen to be consistent with the geometry of the B–DNA helix. Since these three beads are always collinear and their distances are kept constant, based on the number of degrees of freedom we classify this model as a two bead one. The top–down methodology precludes direct transformation of a full–atomistic structure into the CG representation. However, such relationship is unnecessary because the model was not designed to reproduce the results from more detailed methods. The (re)mapping is not required for applying this CG model as long as one is interested solely in the dynamics of DNA hybridization. Here, the fidelity to the 3D structure is rather substituted with an adherence to the 2D hydrogen bond topology. Such bonding network may

[2] There is an earlier version of the model [86] in a four collinear beads variant, with separate beads for base repulsion site and base hydrogen–bonding site.

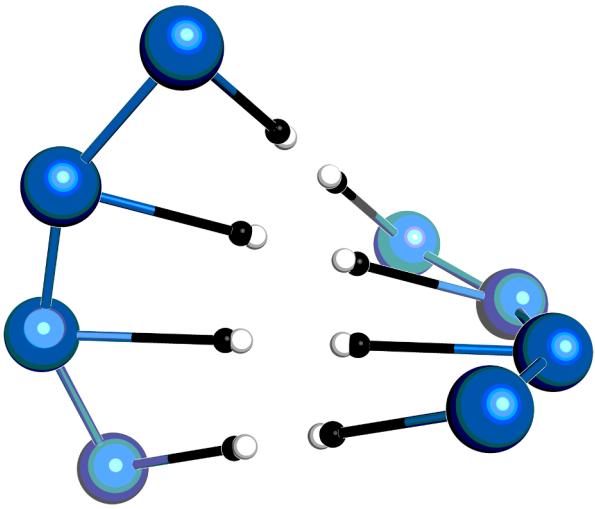

Fig. 6 Four base pair part of a helix in the Ouldridge et al. model [87]. Large beads represent the backbone sites. Small black beads represent the stacking sites and small white beads the base repulsion/hydrogen–bonding sites. In contrast to other presented FFs, this model does not provide a mapping function that links full–atomistic and coarse–grained structures, so the full–atom structure is not shown in the background.

be created by a user or taken from a cadnano program [113], which facilitates the design of DNA Origami.

Presented CG FF is consistent with a general form presented in (2). The potential might be used in both Langevin MD and Virtual Move MC simulation methods [114] (variant of MC simulation by Whitelam et al. to model system dynamics in time). For efficient simulation in the latter one all interactions have to be pairwise, so the authors included in the model only interact between two nucleotides (treated as rigid bodies),

The intrastrand interactions are modeled using three terms:

$$V_{intrastrand} = V_{bond} + V_{stack} + V_{ex} \,, \tag{21}$$

where the V_{bond} term, responsible for the interaction of two backbone beads, uses a finitely extensible nonlinear elastic spring:

$$V_{bond} = -\frac{\varepsilon}{2} \ln\left(1 - \frac{(r - r_0)^2}{\Delta^2}\right) , \tag{22}$$

where r_0 is the equilibrium distance, Δ defines the range of acceptable deviations from the equilibrium (for $r < r_0 - \Delta$ or $r > r_0 + \Delta$ the potential is infinite ∞) and ε reflects the value of the potential on the edges (at $r = r_0 - \Delta$ and $r = r_0 + \Delta$) and controls the steepness of the potential. The stacking term, V_{stack}, is controlled by the Morse potential (see Fig. 2b and (10)) and connects the stacking sites of the base. This term is multiplied by numerous orientation terms that depend on mutual arrangement of bases (see Fig. 7), e.g., preventing the formation of a left–handed helix (see [87, 86] for full equations). Finally, the excluded volume term V_{ex} is responsible for the interactions between the base repulsive site and the neighboring

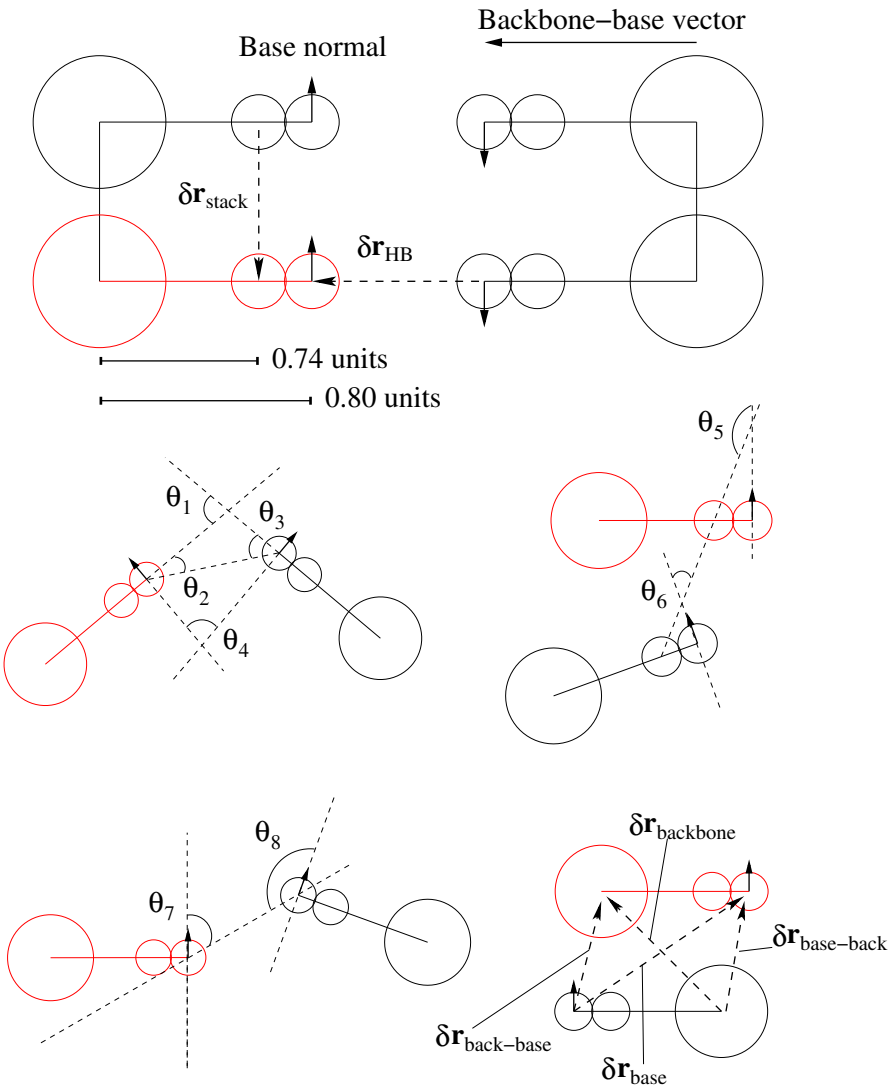

Fig. 7 Topology of interactions presented in the Ouldridge et al. model. The upper part presents the stacking and non–bonded interactions. Middle left, middle right, and bottom left pictures show the angles that modulate the hydrogen bonding and stacking terms. The bottom right figure shows the topology of the excluded volume terms. (Figure was taken from reference [87] and used with permission)

base backbone site and is described by the repulsive part of the Lennard–Jones potential (see Fig. 2a for $r < \sigma$ and (8)).

The interstrand potential is composed of two terms:

$$V_{interstrand} = V_{HB} + V_{cross-stacking}\,, \tag{23}$$

where V_{HB} is the hydrogen bonding and $V_{cross-stacking}$ the cross–stacking potential term. These interactions are calculated between all A and T bases and C and G bases in the system (no secondary structure is supplied), therefore, cutoffs are applied. The V_{HB} term accounts for the interactions of hydrogen bonding sites of two complementary bases and is implemented with the Morse potential (see Fig. 2b and (10) with orientation terms [87, 86] as in the case of V_{stack} (see [87, 86] for full equations). The $V_{cross-stacking}$ term connects the stacking sites of a base and its complementary counterpart neighbor (i.e., *i:j+1* and *i:j-1* interactions). It is implemented with a harmonic potential (see 1a and (4)) multiplied by additional orientation terms [87, 86].

Finally, the non-bonded term, V_{nb}, is an excluded volume potential, which is implemented using the repulsive part of the Lennard–Jones potential (see Fig. 2a and (8) for $r < \sigma$). V_{nb} describes the interactions of the backbone site with the base repulsion site, between the base repulsion sites, and between the backbone sites (but not between the *i:i+1* neighbors).

This FF was applied to simulate the dynamics of DNA tweezers [109] — a DNA system with two arms which can acquire an open or a closed state like real tweezers. The transition between the two states is done by adding two short complementary DNA fragments. These oligomers take part in a sequence of events — hybridization and strand–breaking, but finally they are removed from the system, with the tweezers state altered. The model of Ouldridge et al. [85] was the first CG model applied to DNA tweezers. CG Virtual Move MC simulations [114] helped to understand the free energy changes related to the transition between an open and closed state, caused among others by unfavorable opening up of a second single–stranded region when the displacement begins. This CG model was also applied to simulate a DNA walker [111] in which a short single–stranded DNA fragment moves over a longer strand — a track. The CG Langevin MD simulations pointed to possible problems in this nanostructure, e.g., the authors predicted that in some cases a backward movement of the walker might occur. The CG simulations gave ideas how to avoid this backward movement, e.g., suggested to apply a mechanical tension to the track. The Ouldridge et al. [85] model was also used to simulate kissing loops [115] and Holliday junctions [116] – well-known RNA motifs.

The model of Ouldridge et al. [85] is an interesting approach to modeling nucleic acids — its biggest advantage is a top–down design that sets thermodynamics above structural fidelity. Although the model seems perfect for nanotechnological applications, in the current version it cannot be applied to biological problems. The structural details which are not that important in the nanotechnology field, such as the major and minor groove sizes (which in this model are of the same size), are fundamental for DNA–protein interactions. Also, in the case of RNA, it would be problematic that a starting structure for the CG simulation cannot be supplied from

an external file containing the coordinates of an already folded 3D RNA model and that there is no treatment of non–canonical base pairs.

8 Thermal Denaturation of DNA with a Two–Bead Model

The two-bead per nucleotide model by Drukker et al. [80, 117] was designed to describe the thermal denaturation of DNA. The model was applied in nanomaterial science and used to model DNA translocation in nanopores [118] and in carbon nanotubes [119]. The CG beads are placed in the geometrical center of a backbone group (sugar and phosphate) and a base (see Fig. 5b).

This CG FF uses the standard intrastrand potential scheme (4), where the pseudo–bond V_{bond}, pseudo–angle V_{angle}, and pseudo–dihedral $V_{dihedral}$ potentials are harmonic (see (4), (5), (6) and Fig. 1.) In addition, the *i:i+2* bonds between the backbone beads are added in the intrastrand potential to account for stabilization of the backbone helical conformation since the V_{angle} and $V_{dihedral}$ were insufficient.

In the interstrand potential, this model accounts for the chemical details of hydrogen bonding. The A–T pair is connected by two and the C–G pair by three bonds. Each base can be a donor and/or an acceptor of a hydrogen bond. A and T are both an acceptor and a donor of one bond. G is a donor of two bonds and an acceptor of one. C is a donor of one and an acceptor of two. To assure that interstrand interactions are considered only between the correctly oriented beads, a θ_{ij}^{HB} angle is introduced. This is an angle between a donor backbone, donor base, and acceptor base beads.

$$V_{interstrand}(r,\theta^{HB}) = V_{Morse}(r) - V_{H2}(r) f(\theta^{HB}) \,, \tag{24}$$

$$V_{H2}(r) = \frac{1}{4}(\tanh[\lambda(r-r_2)]-1) \,, \tag{25}$$

$$f(\theta_{ij}^{HB}) = \begin{cases} \frac{1}{2}(\cos(\gamma\theta_{ij}^{HB})+1) & \theta_{min} < \theta_{ij}^{HB} < \theta_{max} \\ 0 & \text{otherwise} \end{cases} . \tag{26}$$

There are three parts of the potential: V_{Morse} is a Morse potential (see (10) and Fig. 2b) that stabilizes a bond between two complementary residues. V_{H2} mimics the solvent effects, which stabilize the denaturated state, and is a switch function (see Fig. 2e for an example of a switch function), with the λ parameter controlling the steepness at the switching distance r_2. Function $f(\theta_{ij})$ describes the effect of the θ_{ij}^{HB} on the total potential (only if θ_{ij}^{HB} is in the range $\theta_{min} - \theta_{max}$). The intrastrand potential can be applied between any two complementary bases so it is not dependent on the inputted secondary structure. The nonbonded potential is implemented using the Lennard–Jones potential (see (8) and Fig. 2a).

This FF was used in 75 ns-long MD simulations and correctly predicted the melting temperatures of 10 base-pair DNA duplexes containing either A–T or C–G pairs [80]. For the A–T and G–C duplexes, the calculated melting temperature error was on average 6.5 K and 18.5 K, respectively. The model was also shown to

give correct melting temperatures of these duplexes containing single mismatches. Introducing a single G–G mismatch to the G–C duplex decreased the melting temperature by 21 K. Such decrease is consistent with the predictions from the thermodynamic models but any quantitative conclusions cannot be made because the thermodynamic models give temperature shifts from 12 to 38 K.

This two–bead FF is useful in the simulations in which the complementary bonds need to be broken such as in DNA melting. With less interaction sites it gives a higher efficiency than three–bead models [68]. This CG FF does not depend on a provided secondary structure as in one–bead models [71]. A two–bead model is the minimal one to be able to introduce base–base orientation terms, as in (24), and this is a necessary condition to determine the presence of a hydrogen bond.

9 One–Bead Model for Linear and Circular DNA

Trovato and Tozzini [71] designed a one–bead model for MD simulations of a linear and circular DNA duplexes and parameterized it to account also for the temperature effects. The nucleotide bead is placed in the position of a phosphorus atom (see Figs. 3 and 5c). This model was also recently modified for RNA helices using an automatic parameterization method based on evolutionary algorithm [66].

The sum of standard terms as in (3) forms the intrastrand potential. The pseudo–bond V_{bond}, pseudo–angle V_{angle}, and pseudo–dihedral $V_{dihedral}$ potentials have the harmonic functional form (see (4), (5), (6) and Fig. 1).

The interstrand potential is added based on the information about the secondary structure. This potential has a specific topology (see Fig. 8). For a complementary pair *i:j* the following pseudo–bonds are created: *i:j*, *i:j+1*, *i+1:j+1*. The term is composed of a Morse with a barrier function (for the graph of the potential function see Fig. 2e):

$$V_{interstrand} = V_{i:j} + V_{i:j+1} + V_{i+1:j+1} \,, \tag{27}$$

$$V_{i:j}(r) = V_0^{i:j}([1 - \exp(-\alpha_{i:j}(r_{kl} - r_0^{i:j}))]^2 - c_{i:j}) sw_{i:j}(r_{i:j}) \,, \tag{28}$$

$$sw_{i:j}(r) = \frac{1}{2} V_1^{i:j}[1 - \tanh(\lambda_{i:j}(r - r_1^{i:j}))] \,, \tag{29}$$

where $V_0^{i:j}$, $r_0^{i:j}$, and $\alpha_{i:j}$ control the shape of the original Morse potential, c affects the energy difference between the energy minimum and unbound state, $\lambda_{i:j}$ controls the slope of the switch function, $V_1^{i:j}$ controls the switch function energy difference, and $r_1^{i:j}$ the position of the switch. Equations (28) and (29) are identical for *i:j+1* and *i+1:j+1*. Though the formula seems complicated it is advantageous; the Morse function (28) makes it possible to account for the breaking of hydrogen bonds and the switch function (29) adds a barrier for long–range electrostatic repulsion.

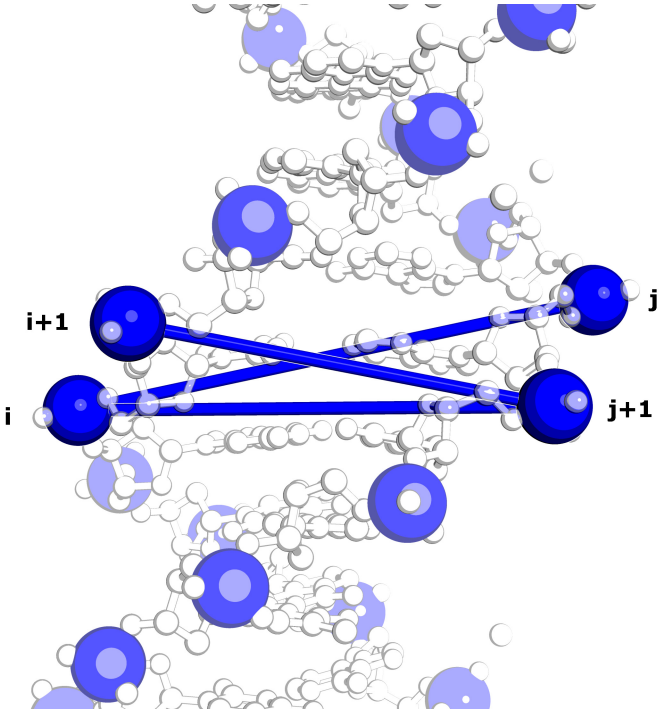

Fig. 8 One–bead representation of DNA. The interstrand interaction topology for a single complementary pair is shown according to the model of Trovato and Tozzini [71].

A similar formula is used for the nonbonded potential:

$$V_{nb}(r) = \begin{array}{l} V_0^{nb}([1-\exp(-\alpha_{nb}(r_{kl}-r_0^{nb}))]^2 - c_{nb}) \\ (1+2sw_2^{nb}(r_{kl}))sw_1^{nb}(r_{kl}) + 2A_{nb}sw_2^{nb}(r_{kl})) \end{array}, \tag{30}$$

$$sw_q^{nb}(r) = \frac{1}{2}V_q^{nb}[1-\tanh(\lambda_q^{nb}(r-r_q^{nb}))]\,, \tag{31}$$

where the A_{nb} parameter controls the addition of a second switch function and thus affects the slope of the "unbound" site of the barrier. Other labels are consistent with (28) and (29), but since two switch functions are used in (30), superscripts in (31) denote the first and the second switch. The authors found that this formula provides for the stabilization of DNA grooves. Since both interstrand and nonbonded potential formulas are computationally expensive, the energy (and force) can be precomputed for a range of distances. Next, their value at a given bead distance, which is between two precomputed points, is interpolated. This procedure saves a lot of time in contrast to calculating exponential and hyperbolic tangents in each simulation step for each pair of beads connected by interstrand or nonbonded interactions.

First, the potential was parameterized based on the potential of mean force, calculated from the experimentally derived 3D structures containing DNA helices. Second, the potential was tuned to match the experimental melting temperatures. The authors validated their CG FF by performing MD simulations of 92 base-pair DNA nano–circles with different twist angles. The effects of the initial twist angle on the nano–circle topology were in agreement with full–atomistic simulations [120, 121]. Next, the authors show the results of CG MD simulations of a DNA plasmid composed of 861 base pairs (approx. $0.3\,\mu$ circumference length) on a microsecond time scale. These MD simulations show that modification of a torsional stress affects the stability of the plasmid and allows forming a denaturation "bubble" [71].

10 One–Bead DNA Model Derived with the Renormalization Group Method

This model of Savalyev et al. [67, 88] is more a parameterization method than a model to study the dynamics of DNA. The authors present a renormalization group optimization method developed by Swendsen [122] and further improved by Lyubartsev and Laaksonen [123], to find the best parameters of a DNA one–bead FF. For the renormalization group method, categorized as the local optimization method, the potential energy function V has to be a linear combination of terms, $V = \sum_i^N k_i * V_i$, with a set of linear combination parameters k_i. In addition, a set of observables S_j that characterize a CG FF has to be defined. These observables must depend on the selected k_i parameters in the potential energy expansion. The aim of the optimization is to find a set of k_i that result in S_j which best resemble the reference data. The observables used by Savalyev et al. were distance distribution, with reference values taken from full–atomistic simulations. In the parameterization procedure one creates a set of k_i parameters and calculates the "susceptibility" of a certain parameter to affect the observables. This susceptibility is expressed as a partial derivative of an S_j observable over a k_i parameter. Next, these derivatives are used to calculate the corrections of parameter sets. This method allows for an objective and effective parameterization, however, it is only applicable to linear combination terms. This means that if the methodology was applied to a harmonic potential $k_i(r - r_0)^2$, it could find an optimal value of the k_i force constant but not the equilibrium distance r_0.

To show the applicability of the renormalization group optimization Savalyev et al. [67] test it on a one-bead CG FF of DNA. In the model the pseudo–atoms are placed in the geometrical center of a nucleotide (see Fig. 5d). The FF uses only pseudo–bond and pseudo–angle terms omitting the pseudo–dihedral term. These terms are a sum of the harmonic, cubic, and quartic terms to include the anharmonicity of bonds (see Fig. 1a):

$$V_{intrastrand} = V_{bond} + V_{angle}\,, \tag{32}$$

$$V_{bond}(r) = k_{r2}(r - r_0)^2 + k_{r3}(r - r_0)^3 + k_{r4}(r - r_0)^4 \,, \tag{33}$$

$$V_{angle}(\theta) = (k_{\theta 2}(\theta - \theta_0)^2 + k_{\theta 3}(\theta - \theta_i)^3 + k_{\theta 4}(\theta - \theta_0)^4 \,. \tag{34}$$

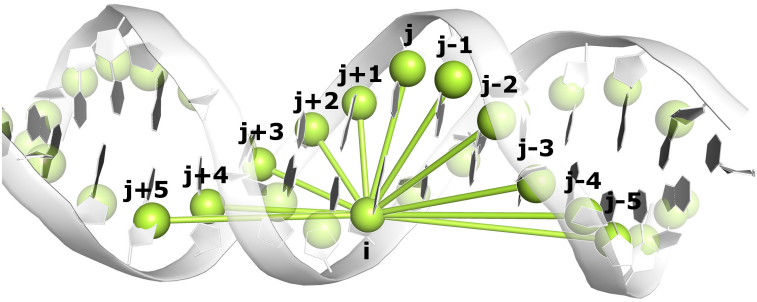

Fig. 9 A DNA helix in a one–bead representation with the beads placed in the geometrical center of each base. The cartoon representation in the background shows the positions of the phosphate backbone (*ribbon*) and bases. The bonds between the beads represent the "fan" interactions, as defined in the Savelyev et. al [67] model. These interactions connect the nucleotide corresponding to bead *i* with eleven nucleotide beads from *j-5* to *j+5* on the complementary strand.

The interstrand terms are implemented using the so-called "fan" interactions. The name originates from their topology (see Fig. 9) because they explicitly connect a nucleotide bead with eleven beads on the opposite strand. Fan interactions are thus *i:j-5* to *i:j+5* interactions in the previously introduced notation (see Fig. 3). These interactions are implemented in the same way as V_{bond} interactions, i.e., as a combination of harmonic, cubic, and quartic terms (see Fig. 1a)

$$V_{fan} = \sum_{-6<m<6} (k_2(r_{i:j+m} - r_0)^2 + k_3(r_{i:j+m} - r_0)^3 + k_4(r_{i:j+m} - r_0)^4) \,. \tag{35}$$

For both the intra– and interstrand potentials, the CG equilibrium distances and angles, as well as the starting values of force constants, are found by matching with the reference distance distributions. The reference distributions are obtained from a 60 ns full–atomistic MD simulations of a 16 base-pair DNA duplex performed with the AMBER parmbsc0 FF [124]. The CG force constants are optimized based on the difference of distance distribution from 20 ns CG MD simulations (preceded by a 5 ns heating and 10 ns equilibration) and from the reference full–atomistic MD simulation.

In the first Savalyev et al. paper [67] the nonbonded interactions were modeled using the following equation to match the potential of mean force:

$$V_{el}(r) = A\frac{\exp -r/k_D}{r^4} + \frac{q_i q_j \exp(-(r - r_{bead})/k_D)}{4\pi\varepsilon_0\varepsilon_w r(1 + r_{bead}/k_D)} \,. \tag{36}$$

In the next publication [88] of this group, another expression for this potential was found, which better describes the interaction of DNA beads with ions:

$$V_{el}(r) = \frac{A}{r^{12}} + \sum_{k=1}^{3} B_k \exp(-C_k(r - R_k)) + \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_w r}\,, \tag{37}$$

where A and B_k, C_k, and R_k are adjustable parameters.

According to the authors, the optimized potential results in a radial distribution function that is consistent with the full–atomistic simulation. Though the calculated DNA persistence length is overestimated nearly two times, it is in agreement with the results from full–atomistic model used for parameterization [125]. If the force constants are rescaled by a factor of 0.7, the DNA persistence length differs by less than 10% from the experiments. This value is much lower than the one obtained by Knotts et al. [68] for a three–bead model (error higher than 200% was considered good enough in this work). However, Knotts et al. modeled a much longer 1489 base-pair DNA and Savalyev et al. tested less than 100 base pairs. Also, Knotts et al. tested a larger variety of DNA chain lengths and sequences. The difference in the persistence length in the model of Savalyev et al. is attributed to widening of the "fan" interactions allowing for larger internal fluctuations. Savalyev et al. estimated also the effects of NaCl salt concentration on supercoiling of a 90–base pair DNA plasmid but made only qualitative conclusions.

This work is a success of automatic optimization methods for CG FFs. By performing a systematic procedure, the authors obtained a reasonable one–bead FF for DNA. However, in comparison with the other DNA FFs (e.g., Trovato and Tozzini [71] described below), the model of Savalyev et al. [67] contains too many potential terms and parameters. The latter model requires 10 quartic, cubic, and harmonic terms per one nucleotide pair and the model of Trovato and Tozzini [71] requires only three Morse with a barrier terms. Since the Savalyev et al. [67] algorithm cannot cope with non-linear parameters, in order to find an efficient DNA model, a perfect strategy would be to add new terms. However, this would result in a much slower performance of the method.

11 One–Bead Model for RNA Structure Prediction from Tertiary Contacts

The Nucleic Acid Simulation Tool (NAST) by Jonikas et al. [79] was designed to generate candidate tertiary structures of RNAs in order to solve the RNA structure prediction problem. This is the only FF presented in this chapter that is not intended to analyze the internal motions of nucleic acids. MD is only used as a means of sampling the conformational space. The generated trajectories do not serve to understand the RNA folding process. Only the final 3D RNA structure is of value. Also, NAST was not designed to perform the tertiary structure prediction from scratch, like the model by Ding et al. [69]. In advance, one has to provide the RNA secondary structure (from a secondary structure prediction software [23])

and information about a few tertiary contacts (from chemical or spectroscopic methods [126, 127, 128, 129]). Though NAST imposes these a priori requirements, it is useful. For example, for RNAs which are difficult to crystalize but were preliminary studied with some chemical methods, NAST can quickly perform a wide search of possible conformations and generate multiple candidate structures for further studies. The quality of these structures is assessed by comparing them with pairwise distance distributions from small angle X–ray scattering experiments, solvent accessibility data, and the NAST energy tool. The RNA structures that score best can be later tested with full–atomistic models.

In this model RNA is represented using a single bead, centered on the C3' atom of the sugar group (see Figs. 4c and 5e). The total potential energy of RNA is a sum of four terms:

$$V_{total} = V_{intrastrand} + V_{interstrand} + V_{tertiary} + V_{nb}, \tag{38}$$

where $V_{intrastrand}$, $V_{interstrand}$, and V_{nb} are consistent with (2) and $V_{tertiary}$ is a restraint for tertiary RNA contacts.

The intrastrand potential is a sum of three terms previously shown in (3) with the pseudo–bond V_{bond} and pseudo–angle V_{angle} potentials using the harmonic functions (see (4), (5) and Fig. 1a,b) but the V_{angle} term uses two different force constants, k_θ, for single– and double–stranded regions. The pseudo–dihedral potential $V_{dihedral}$ is implemented using a cosine potential (see (7) and Fig. 1c).

The interstrand potential is composed in a similar manner as the intrastrand one of bonded (V_{bond}, (4)), angle (V_{angle}, (5)), and dihedral ($V_{dihedral}$, (7)) terms. The pseudo–bonds connect only complementary base pairs, the pseudo–angle is between a complementary bond and the next neighbor on the first strand, and pseudo–dihedrals are the following: *j-1:j:i:i+1* and *j+1:i-1:i:i+1* (for details of the *i,j* notation see Fig. 3).

The nonbonded interactions, V_{nb}, are implemented using a repulsive–only potential

$$V_{rep}(r) = 4V_0 \left(\frac{\sigma}{r}\right)^{12}. \tag{39}$$

The user-supplied tertiary interactions ($V_{tertiary}$) are implemented as pseudo–bonds with the assumption that the nucleotides are close to each other in the final structure (see (4)). If a particular tertiary contact is uncertain, the authors recommend using a much smaller force constant for that contact.

NAST is a knowledge–based model. It was parameterized by fitting the presented potential functions to the Boltzmann inversion of distance distributions from three high-resolution ribosome structures. The authors test 3D structure predictions for tRNA and a P4–P6 medium–sized RNA with the root mean square deviations from the crystal structures equal to 8 Å and 16.3 Å, respectively. Another measure, the GDS–TS score (the average percentage of residues that are within 1, 2, 4, and 8 Å of their reference position), was equal to 0.2 and 0.06. These numbers are larger than the ones obtained by Ding et al. [69], where the root mean square deviation was on average less than 4 Å. This difference is accounted to the smaller number of details

that can be captured in a one–bead NAST FF in comparison with the three–bead FF of Ding et al. [69].

NAST relies on the provided secondary structure and tertiary contacts. The outputted 3D prediction is the best one that reflects the applied constraints. In a more complex model of Ding et al. [69], in which the mutual base orientation is present, the tertiary interactions can be predicted. However, the predictions of the NAST one–bead model are useful because they still give a lower deviation from the crystal structures than a random structure of the same sequence and a similar radius of gyration. This model can be useful, e.g., to rebuild missing loops.

NAST also provides a supplementary tool, C2A, to rebuild a full–atomistic structure from a CG model [130]. The remapping is performed by finding short fragments of a matching sequence in the 3D structure of the ribosome and then building and optimizing the user's RNA structure. However, remapping is as good as the given CG model. Two measures were used to asses the quality of remapping — root mean square deviations and interaction network fidelity (INF) score, which measures the number of correctly found base pairs and stacking [131]. If a tRNA [78] PDB structure is reduced to a CG representation and then used as a template for C2A rebuilding the root mean square deviation of 2.81 Å and the INF score of 69% are achieved. However, if the best NAST-predicted tRNA structure is used, the root mean square deviation becomes 8.30 Å and the INF score drops to 46% (35% if only base pairing is taken into account in the INF score).

12 One–Bead Model for Large RNAs

Contrary to the previous NAST model, this one was designed for large RNAs and ribonucleoprotein particles, namely the 30S ribosomal subunit which contains over 1500 nucleotide long RNA chain (16S RNA). The model is based on previous ones for supercoiled DNA [132] and ribosomal RNA [72, 73, 74] and can be classified as either one bead or one and a half bead per residue model. The residues are represented as beads centered on a P atom (for nucleic acids) or C_α atom (for proteins). In order to achieve the correct helical conformation, additional space–filling dummy beads ("X-atoms") are placed in the geometrical center of complementary base pairs (apart from the last base pair in the helix, see Figs. 5f and 10).

The intrastrand potential is a sum of standard terms shown in equation 3 with V_{bond}, V_{angle}, and $V_{dihedral}$ calculated with harmonic functions ((4), (5), (6), and Fig. 1a,b,c). The pseudo–bond and pseudo–angle force constants are higher in helical regions than in non-helical ones.

The interstrand potential is also composed of pseudo–bonded, pseudo–angle, and pseudo–dihedral terms with the same definitions of potentials as in (4), (5), and (6). Here, the pseudo–bonds connect the nucleotide beads with the corresponding X-atoms and complementary bases (*i:j+1* and *i:j-1* configurations). The pseudo–angle interactions are present in the P–X–P configuration along a complementary bond. There is also a dihedral angle connecting *i-1:i:j:j+1* (see Fig. 10). Proteins are modeled as simple elastic networks where the harmonic pseudo–bonds are created between all protein beads that are closer than 8 Å to each other.

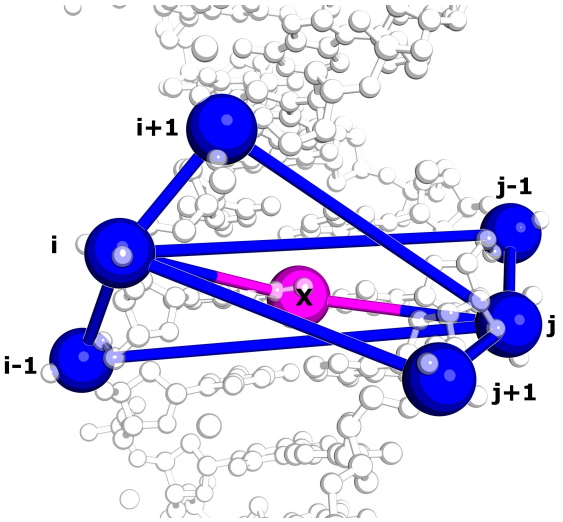

Fig. 10 Topology of interstrand bonds in the CG model of Malhotra et al. [72, 73, 74] showing the placement of beads on the phosphorus P atoms. The central nucleotide pair is represented with two pseudo–atoms: in the position of the P atom and the dummy X atom in the middle. Two neighboring base pairs are presented only using P atoms. There are 6 pseudo–bonds associated with this i:j pair (two P–X in the middle and four P–P bonds) and a pseudo–angle P–X–P.

The nonbonded potential consists of restraints on the helix–helix and protein–RNA distances (V_{rest}) and a volume exclusion term (V_{excl}) among P-, X-, and C-atoms

$$V_{nb} = V_{rest} + V_{excl}\ , \tag{40}$$

$$V_{restr}(r) = \begin{cases} k_2(r-r_2)^2 & r < r_2 \\ 0 & r_2 < r < r_3 \\ k_3(r-r_3)^2 & r_3 < r < r_4 \\ k_3\left[\frac{b}{r-r_3}+1\right] & r > r_4 \end{cases}\ , \tag{41}$$

$$V_{excl}(r) = K_{excl}(r-d_0)^2 \quad r < d_0\ . \tag{42}$$

For graphical representation of distances r_2, r_3, r_4 see Fig. 2f. $a = 3(r_4 - r_3)^2$ and $b = -2(r_4 - r_3)^3$. k_3 is a constant describing the steepness of the restraint potential for $r > r_3$. The V_{restr} restraints, presented in (41), are applied to all P–C_α pairs that lie within a 10 Å cutoff in the reference structure. For helices, these restraints are also applied to non-canonical base pairs (i.e. nucleotides hydrogen bonded with other than Watson–Crick type bonding). However, the ones explicitly enumerated by Wimberly et al. [133], in this chapter describing the crystal structure of *Thermus thermoplius* 30S ribosomal subunit, are considered in the interstrand potential on the same basis as Watson–Crick ones. Other ones, i. e., all P–P pairs within a 6 Å cutoff distance that are not already connected, are included in the restraints term. This term provides a certain freedom of movement between the r_2 and r_3 distances (which are independently set for each type of atom pairs), however, the movement

is penalized when going outside of this range (see Fig. 2f). Therefore, this restraint term generates a bias toward a starting structure. The space exclusion term, V_{excl}, prohibits two nucleotide beads from getting closer to each other than d_0.

This is also a knowledge–based potential. The crucial parameters for the model, i.e., the parameters of protein–RNA distance restraints are taken from the high resolution *Thermus thermophilus* 30S ribosome subunit structure. Other parameters are taken from the lower resolution ribosome models and/or older models [73]. The force constants k_2 and k_3 are optimized to maintain the crystal structure of the 30S ribosome subunit at room temperature, while allowing the flexibility of the free 16S ribosomal RNA.

This model was designed and applied to study the assembly of proteins to 16S RNA of the small ribosomal subunit. Stagg et al. [134] explored one of the assembly paths using the MC simulated annealing technique. The starting model of 16S RNA contained only the information on its secondary structure. The restraints of (41) guided the ribosomal proteins from the initial random positions to their appropriate binding sites on 16S RNA. The authors examined the change in the fluctuations of 16S RNA upon binding of proteins and predicted the contributions of each protein to the organization of its binding site. Cui et al. [74] also used this model to investigate the assembly of ribosomal proteins but applied MD simulations and additionally studied the flexibility of 16S RNA during adding the proteins at various orders. The experimental assembly paths were reproduced even with such a simple CG model.

13 One–Bead Model for Protein-RNA Complexes

This model was developed to perform MD simulations of macromolecular complexes of proteins and RNA on microsecond time scales. In the original publication it was applied to investigate the flexibility of the whole ribosome [75]. In this model a single bead represents a nucleotide (centered on a phosphorus atom, see Fig. 5c) or an amino acid (centered on a C_α atom).

The residues of the backbone are connected with the intrastrand harmonic potential which is a sum of the pseudo–bond (4), pseudo–angle (5), and pseudo–dihedral (6) terms as in equation 3. The classical Morse potential was also tested for the intrastrand terms but since these terms connect the residues that are no more than four CG beads apart the authors found that harmonic functions are sufficient.

The interstrand $V_{interstrand}$ energy term is based on an externally provided secondary structure for RNA and uses a harmonic function (see (4) and Fig. 1a). This potential accounts for the canonical hydrogen bonds that appear in the RNA motifs.

The nonbonded potential is implemented using Morse functions and its general form is:

$$V_{nb}(r_{ij}) = A_{P,C_\alpha}(r_0^{ij})[1 - \exp(-\alpha(r_{ij} - r_0^{ij}))]^2 \,. \tag{43}$$

The strength of this potential is adjusted by the $A_{P,C_\alpha}(r_0^{ij}) = a\exp(-r_0^{ij}/b)$ function. The constants a and b are based on the interacting bead types (different for P and C_α). For local short-range interactions (within a predefined cut–off of 12Å for C_α and 20Å for P pairs), the r_0^{ij} equilibrium values are taken from the starting structure.

For all the other long-range nonbonded interactions beyond the short-range cutoff (but within a certain limits), r_0^{ij} assumes three different values for P–P, C_α–C_α, and C_α–P pairs and does not depend on the starting conformation. Therefore, the model is only locally biased toward the starting structure even though the possibility of breaking of the interactions exists also for the nonbonded short-range contacts.

Overall, the model is an extension of an elastic network model but since the nonbonded interactions are represented with the Morse potential it allows for larger fluctuations from the initial conformation than the harmonic potential. The model was parameterized based on the Boltzmann inversion procedure with the distribution functions taken from a single ribosome structure so it is not immediately transferable to other systems. This CG FF was used to perform half a microsecond MD simulations of the ribosome and determine global collective motions of the ribosome fragments. The correlations of motions showed the possible movements during the translocation of tRNAs on the ribosome. The movement of the distant ribosomal stalks, positioned on the opposite sides of the tRNA path, appeared to be coupled with the ratchet-like motion of the subunits.

14 One–Bead Model for Protein–DNA Complexes

Later a similar anharmonic elastic network methodology was applied in MD simulations of the nucleosome [76]. The nucleosome is a basic unit of chromatin and is composed of double-stranded DNA wrapped around histone proteins.

The interstrand and nonbonded functional terms are similar as in the model of Trylska et al. [75]. However, in order to account for the helicity of the histone proteins and DNA, the nucleosome model required slightly different formulation of the intrastrand potential:

$$V_{intrastrand} = V_{1-2} + V_{1-3} + V_{1-4} + V_{1-5}\,, \tag{44}$$

where V_{1-n} terms are implemented using a harmonic potential (see (4) and Fig. 1a). For the α–helical regions of the proteins, all terms in (44) are included. However, in unstructured regions or loops only V_{1-2} and V_{1-3} are included, whereas V_{1-4} and V_{1-5} are modeled as nonbonded interactions. In the case of DNA beads V_{1-5} is not required.

The model was parameterized with the Boltzmann inversion procedure based on short 50 ns full-atomistic MD simulations of the nucleosome [76]. Next, it was applied to perform multiple 10 microsecond scale MD simulations of the nucleosome complex [135]. In these simulations a biologically relevant partial unwrapping of the DNA from the nucleosome core was observed. Further remapping to all-atom model provided a better insight into the interactions that are formed by histone tails after the DNA detachment from the nucleosome core. One of the histone tails (H3) was seen to stabilize the nucleosome in the open state by interacting with the nucleosome core. The removal of this H3 tail in the simulations precluded the formation of such a long-lived detachment of the DNA terminal segment from the nucleosome protein core. This suggests an active role of this tail not only in the detachment

of the DNA end from the nucleosome core but also in preventing the nucleosomal DNA from rewrapping.

15 Conclusions

Residue resolution FFs may be applied to solve various kinds of problems in the nucleic acid field, ranging from RNA structure prediction to global motions of large ribonucleoprotein complexes. We have presented a limited set of CG FFs, with the number of beads ranging from one to three per nucleotide. Even in this bead range the design and applicability of the FFs differ. In one bead models the interaction network is based on an externally supplied secondary structure or native contacts from a reference structure. Adding a second bead allows for the secondary structure to be dynamically modified because the orientation of an interstrand bond with regard to the backbone can be measured. Overall, increasing the number of beads corresponds to removing the bias from the system. On the other hand, if one accepts the limitations of one–bead models, problems on much larger spatial and temporal scales may be investigated. For example, the Jonikas et al. [79] one-bead model was easily applied to a 158–nucleotide structure but the three-bead Ding et al. [69] model only to RNA chains shorter than 100 nucleotides. Also, the CG FFs used for large macromolecular complexes, such as the nucleosome or ribosome, are one–bead FFs.

There are two other crucial things to consider when choosing one- to three-bead models. First, with one bead models it is problematic to achieve a correct helical twist. Creating bonds only between complementary pairs, which is easily applied in two- or three-bead models, is not sufficient to keep the helicity in one-bead models. The remedy is to create dummy atoms in the middle of a helix (as in the model of Cui et al. [74]), provide multiple pseudo–bonds per single complementary pair (Savelyev et al. [67], Trovato et al. [71]) or use multi–body terms — angle and dihedral over the interstrand bonds (Jonikas et al. [79]). These ”tricks” were not required, in the model of Trylska et al. [75] because to stabilize the helical structure the equilibrium distances were taken from the native structure. Adding the terms that ensure the correct helicity may give reasonable dynamics but requires higher computational time. Second thing to consider is that neither of one-bead models applies interaction terms that are nucleotide-specific[3]. Even if such interactions were implemented, they would be inefficient since there is no information about the relative orientation of bases. The two- and three-bead models easily incorporate the base specificity.

There are also residue-resolution nucleic acid models with more than three beads per nucleotide, so one may ask if it is worth going beyond the FFs presented in this chapter. The four- or more bead per nucleotide models include more details such as base dipole moments [47] or non–canonical hydrogen bonding schemes [44]. Niewieczerzał et al. [136] compared three CG models with different number of beads per nucleotide: two, three [68], and four/five (depending on the nucleotide

[3] Some of one–bead models, e.g., Trovato et al. [71], assign a mass consistent with the base type in MD simulations but it has a limited effect on the interactions.

type). All three models were applied to a problem of mechanical stretching and twisting of the DNA duplexes. The authors show that the number of beads does not affect the mechanical properties of DNA at low and moderate temperatures, but may become an issue at room temperature.

When comparing the three-bead CG models we also have to consider their applicability to other tasks than these for which they were designed. Typically, their target is narrow and CG FFs are not transferable to other problems or systems. For example, the Hyeon et al. [84] potential was created to answer a specific question about a particular RNA hairpin. A desirable CG FF would be the one that could be easily applied to different sets of problems, i.e., a FF with a clear parameterization procedure and a universal formulation of the potential energy function. A good example is the model of Knotts et al. [68] since this model can be easily implemented and modified. This task would be more difficult with the model of Ding et al. [69]. Despite promising results for the RNA structure prediction, its applicability and possibilities for modifications are limited because its formulation using a non–standard engine, discrete MD, makes this potential much harder to re-implement. There are multiple codes available to provide classical MD or MC procedures, and to use the model of Ding et al. [69] one would have to rely on the authors' or own in-house made code. The model of Hyeon et al. [84] was tuned for a particular molecule, however, the authors show the source of parameterization parameters and it should be easy to re-implement the model for a different task. Another good example of an extendable model is the one designed by Trylska et al. [75]. It was originally created and parameterized for a particular complex — the ribosome. However, there are other studies that applied this model for a large system involving long chains of DNA, not RNA — the nucleosome [76, 135]. The model is also implemented in a freely available software RedMD [137] (http://bionano.cent.uw.edu.pl/Software).

The transferability of the present CG models is insufficient and new models will certainly be needed for particular applications. However, future efforts have to be also put to solve methodological problems. Just to mention two of such problems: the definition of the reference state in the Boltzmann inversion procedure and generalization of simulation results obtained for isolated, small systems to larger volumes. In the first problem we go back to (1), where a function $d_0(r)$ has been introduced as a reference state. The FF parameters depend on this function and its choice is often arbitrary. The second problem, mentioned by Ouldridge et al.[138, 87], refers to the fact that typically CG simulations are performed in small volumes with only a single set of interacting molecules. The process of single DNA duplex formation may give different melting temperatures than when using many duplexes in a larger volume. The solutions to extrapolate the results of a small-size simulation to a larger one have been proposed [138, 87].

There is still room for improvement in the field of low–resolution nucleic acid models. For example, creating an unbiased model of the ribosome is still an open problem and it would provide better insight into the mechanics of this system in comparison with the model based on the concept of native contacts. There is also a need to create more formal protocols for the parameterization of CG FFs and assessment of the quality of parameters. Unfortunately, most authors are vague about

the parameterization details. In some parameterizations there is no account of how well the chosen potential was fitted to experimental data (by means of for example the R^2 regression parameter). The correctness of the model is proven only by simulations of selected test cases but more details on the parameterization would give better confidence in these models. Another issue is that most authors do not give hard evidence why a certain potential energy functional term was used. Test cases that would justify the use of a particular potential form would be of great value. A good remedy for the parameterization problems might be the use of automated procedures to derive the parametes, like the one mentioned by Savelyev et al. [67] using renormalization group approach or developed by us [66] implementing the evolutionary algorithm.

Acknowledgements. The authors acknowledge support from University of Warsaw (ICM/G31-4 and CeNT/BST), Polish Ministry of Science and Higher Education (N N301 245236), National Science Centre (DEC-2011/03/N/NZ2/02482 and DEC-2012/05/B/NZ1/00035), Foundation for Polish Science (TEAM/2009-3/8 project co-financed by European Regional Development Fund operated within Innovative Economy Operational Programme).

References

1. Venter, J.C., et al.: Science 291(5507), 1304 (2001)
2. International Human Genome Sequencing Consortium. Nature 409(6822), 860 (2001)
3. Flicek, P., et al.: Nucleic Acids Res. 39(Database Issue), D800 (2011)
4. Vinograd, J., Lebowitz, J., Radloff, R., Watson, R., Laipis, P.: Proc. Natl. Acad. Sci. USA 53(5), 1104 (1965)
5. Wang, J., Peck, L., Becherer, K.: Cold Spring Harbor Symposia on Quantitative Biology 47, 85 (1983)
6. Mizushima, T., Kataoka, K., Ogata, Y., Inoue, R.I., Sekimizu, K.: Mol. Microbiol. 23(2), 381 (1997)
7. Mizushima, T., Natori, S., Sekimizu, K.: Mol. Gen. Genet. 238(1-2), 1 (1993)
8. Choi, C.H., Kalosakas, G., Rasmussen, K.O., Hiromura, M., Bishop, A.R., Usheva, A.: Nucleic Acids Res. 32(4), 1584 (2004)
9. Richmond, T.J., Davey, C.A.: Nature 423(6936), 145 (2003)
10. Mergny, J.L., Lacroix, L.: Oligonucleotides 13(6), 515 (2003)
11. Mattick, J.S., Makunin, I.V.: Human Molecular Genetics 15(Spec No), R17 (2006)
12. Al-Hashimi, H.M., Walter, N.G.: Curr. Opin. Struct. Biol. 18(3), 321 (2008)
13. Brion, P., Westhof, E.: Annu. Rev. Biophys. Biomol. Struct. 26, 113 (1997)
14. Leontis, N.B., Westhof, E.: Curr. Opin. Struct. Biol. 13(3), 300 (2003)
15. Capriotti, E., Renom, M.M.: BMC Bioinformatics 11(1), 322 (2010)
16. Bloomfield, V.A., Crothers, D.M., Tinoco, I.J.: Nucleic acids: structures, properties and functions, 1st edn. University Science Books (2000)
17. Tucker, B.J., Breaker, R.R.: Curr. Opin. Struct. Biol. 15(3), 342 (2005)
18. Narberhaus, F., Waldminghaus, T., Chowdhury, S.: FEMS Microbiol. Rev. 30(1), 3 (2006)
19. Berg, J.M., Tymoczko, J.L., Stryer, L.: Biochemistry, 7th edn. W.H. Freeman (2010)
20. Lu, Z.J., Turner, D.H., Mathews, D.H.: Nucleic Acids Res. 34(17), 4912 (2006)
21. Turner, D.H., Mathews, D.H.: Nucleic Acids Res. 38(Database Issue), D280 (2010)

22. Kibbe, W.A.: Nucleic Acids Res. 35(suppl. 2), W43 (2007)
23. Mathews, D.H., Turner, D.H.: Curr. Opin. Struct. Biol. 16(3), 270 (2006)
24. Rother, K., Rother, M., Boniecki, M., Puton, T., Bujnicki, J.M.: J. Mol. Model., 2325–2336 (2011)
25. Bernstein, F.C., et al.: Arch. Biochem. Biophys. 185(2), 584 (1978)
26. Leach, A.: Molecular Modelling: Principles and Applications, 2nd edn. Prentice-Hall (2001)
27. Skolnick, J., Koliński, A.: Science 250(4984), 1121 (1990)
28. Koliński, A., Skolnick, J.: Proteins 18(4), 338 (1994)
29. Ma, J.: Structure 13(3), 373 (2005)
30. McCammon, J.A., Gelin, B.R., Karplus, M.: Nature 267(5612), 585 (1977)
31. Schlick, T.: Molecular Modeling and Simulation: An Interdisciplinary Guide (Interdisciplinary Applied Mathematics), 2nd edn. Springer (2010)
32. Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B., Woods, R.J.: J. Comput. Chem. 26(16), 1668 (2005)
33. Cheatham, T.E., Young, M.A.: Biopolymers 56(4), 232 (2000)
34. Brooks, B.R., et al.: J. Comput. Chem. 30(10), 1545 (2009)
35. MacKerell, A.D., Banavali, N., Foloppe, N.: Biopolymers 56(4), 257 (2000)
36. Shaw, D.E., Dror, R.O., Salmon, J.K., et al.: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, SC 2009, pp. 39:1–39:11. ACM, New York (2009)
37. Shaw, D.E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R.O., et al.: Science 330(6002), 341 (2010)
38. Freddolino, P.L., Liu, F., Gruebele, M., Schulten, K.: Biophys. J. 94(10), L75 (2008)
39. Guvench, O., Brooks, C.L.: J. Comput. Chem. 25(8), 1005 (2004)
40. Forrey, C., Muthukumar, M.: Biophys. J. 91(1), 25 (2006)
41. Xia, Z., Gardner, D.P., Gutell, R.R., Ren, P.: J. Phys. Chem. B 114(42), 13497 (2010)
42. Poulain, P., Saladin, A., Hartmann, B., Prévost, C.: J. Comp. Chem. 29(15), 2582 (2008)
43. Dans, P.D., Zeida, A., Machado, M.R., Pantano, S.: J. Chem. Theory Comp. 6(5), 1711 (2010)
44. Pasquali, S., Derreumaux, P.: J. Phys. Chem. B 114(37), 11957 (2010)
45. Rudnicki, W.R., Bakalarski, G., Lesyng, B.: J. Biomol. Struct. Dyn. 17(6), 1097 (2000)
46. Maciejczyk, M., Rudnicki, W.R., Lesyng, B.: J. Biomol. Struct. Dyn. 17(6), 1109 (2000)
47. Maciejczyk, M., Spasic, A., Liwo, A., Scheraga, H.A.: J. Comp. Chem. 31(8), 1644 (2010)
48. Cieplak, M., Sułkowska, J.I.: Multiscale approaches to protein modeling: structure prediction, dynamics, thermodynamics and macromolecular assemblies. In: Koliński, A. (ed.), ch. 8, pp. 179–208. Springer (2010)
49. Arya, G., Zhang, Q., Schlick, T.: Biophys. J. 91(1), 133 (2006)
50. Bruant, N., Flatters, D., Lavery, R., Genest, D.: Biophys. J. 77(5), 2366 (1999)
51. Jian, H., Schlick, T., Vologodskii, A.: J. Mol. Biol. 284(2), 287 (1998)
52. Allison, S.A., McCammon, J.A.: Biopolymers 23(2), 363 (1984)
53. Olson, W.K., Zhurkin, V.B.: Curr. Opin. Struct. Biol. 10(3), 286 (2000)
54. Morriss-Andrews, A., Rottler, J., Plotkin, S.S.: J. Chem. Phys. 132(3), 30 (2010)
55. Mergell, B., Ejtehadi, M.R., Everaers, R.: Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys. 68(2 Pt. 1), 15 (2003)
56. Olson, W.K.: Macromolecules 8, 272 (1975)
57. Olson, W.K.: Biopolymers 15, 859 (1976)
58. Olson, W.K.: Biopolymers 18(5), 1213 (1979)

59. Vorobjev, Y.N.: Biopolymers 29(12-13), 1503 (1990)
60. Liwo, A., Czaplewski, C., Oldziej, S., Rojas, A., Kazmierkiewicz, R., Makowski, M., Murarka, R., Scheraga, H.: In: Voth, G. (ed.) Coarse-Graining of Condensed Phase and Biomolecular Systems, ch.8, pp. 107–122. Taylor & Francis (2008)
61. Liwo, A., He, Y., Scheraga, H.A.: Phys. Chem. Chem. Phys. 13(38), 16890 (2011)
62. Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.H., Srinivasan, A.R., Schneider, B.: Biophys. J. 63(3), 751 (1992)
63. Reith, D., Pütz, M., Müller-Plathe, F.: J. Comput. Chem. 24(13), 1624 (2003)
64. Reith, D.: Comput. Phys. Commun. 148(3), 299 (2002)
65. Hülsmann, M., Köddermann, T., Vrabec, J., Reith, D.: Comput. Phys. Commun. 181(3), 499 (2010)
66. Leonarski, F., Trovato, F., Tozzini, V., Trylska, J.: Proceedings of the 9th European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, EvoBIO 2011, pp. 147–152. Springer, Heidelberg (2011)
67. Savelyev, A., Papoian, G.A.: Biophys. J. 96(10), 4044 (2009)
68. Knotts, T.A., Rathore, N., Schwartz, D.C., De Pablo, J.J.: J. Chem. Phys. 126(8), 084901 (2007)
69. Ding, F., Sharma, S., Chalasani, P., Demidov, V.V., Broude, N.E., Dokholyan, N.V.: RNA 14(6), 1164 (2008)
70. Ding, D., Dokholyan, N.V.: Trends Biotechnol. 23, 450 (2005)
71. Trovato, F., Tozzini, V.: J. Phys. Chem. B 112(42), 13197 (2008)
72. Malhotra, A., Tan, R.K., Harvey, S.C.: Biophys. J. 66(6), 1777 (1994)
73. Malhotra, A., Harvey, S.C.: J. Mol. Biol. 240(4), 308 (1994)
74. Cui, Q., Tan, R.K.Z., Harvey, S.C., Case, D.A.: Multiscale Model. Simul. 5(4), 1248 (2006)
75. Trylska, J., Tozzini, V., McCammon, J.A.: Biophys. J. 89(3), 1455 (2005)
76. Voltz, K., Trylska, J., Tozzini, V., Kurkal-Siebert, V., Langowski, J., Smith, J.: J. Comput. Chem. 29(9), 1429 (2008)
77. Kolk, M.H., Heus, H.A., Hilbers, C.W.: EMBO J. 16(12), 3685 (1997)
78. Sussman, J.L., Holbrook, S.R., Warrant, R.W., Church, G.M., Kim, S.H.: J. Mol. Biol. 123(4), 607 (1978)
79. Jonikas, M.A., Radmer, R.J., Laederach, A., Das, R., Pearlman, S., Herschlag, D., Altman, R.B.: RNA 15(2), 189 (2009)
80. Drukker, K., Schatz, G.C.: J. Phys. Chem. B 104(26), 6108 (2000)
81. DeMille, R.C., Cheatham, T.E., Molinero, V.: J. Phys. Chem. B 115(1), 132 (2011)
82. Freeman, G.S., Hinckley, D.M., De Pablo, J.J.: J. Chem. Phys. 135(16), 165104 (2011)
83. Prytkova, T.R., Eryazici, I., Stepp, B., Nguyen, S.B., Schatz, G.C.: J. Phys. Chem. B 114(8), 2627 (2010)
84. Hyeon, C., Thirumalai, D.: Proc. Natl. Acad. Sci. USA 102(19), 6789 (2005)
85. Ouldridge, T.E., Louis, A.A., Doye, J.P.K.: Phys. Rev. Lett. 104(17), 4 (2009)
86. Ouldridge, T.E., Louis, A.A., Doye, J.P.K.: J. Chem. Phys. 134(8), 085101 (2010)
87. Ouldridge, T. (ed.): Coarse-Grained Modelling of DNA and DNA Self-Assembly. Springer, Heidelberg (2012)
88. Savelyev, A., Papoian, G.A.: Proc. Natl. Acad. Sci. USA 107(47), 20340 (2010)
89. Hoang, T.X., Cieplak, M.: J. Chem. Phys. 112, 6851 (2000)
90. Swendsen, R.H., Wang, J.S.: Phys. Rev. Lett. 57, 2607 (1986)
91. Kumar, S., Bouzida, D., Swendsen, R.H., Kollman, P.A., Rosenberg, J.M.: J. Comput. Chem. 13, 1011 (1992)
92. Sambriski, E.J., Schwartz, D.C., De Pablo, J.J.: Biophys. J. 96(5), 1675 (2009)
93. DeMille, R.C., Molinero, V.: J. Chem. Phys. 131(3), 034107 (2009)

94. Chen, Y., Ding, F., Nie, H., et al.: Arch. Biochem. Biophys. 469, 4 (2008)
95. Mathews, D.H., Sabina, J., Zuker, M., Turner, D.H.: J. Mol. Biol. 288, 911 (1999)
96. Zuker, M.: Nucleic Acids Res. 31(13), 3406 (2003)
97. Sharma, S., Ding, F., Dokholyan, N.V.: Bioinformatics 24(17), 1951 (2008)
98. Klimov, D.K., Thirumalai, D.: Proc. Natl. Acad. Sci. USA 97, 7254 (2000)
99. Rüdisser, S., Tinoco, I.: J. Mol. Biol. 295(5), 1211 (2000)
100. Liphardt, J., Onoa, B., Smith, S.B., Tinoco, I., Bustamante, C.: Science 292(5517), 733 (2001)
101. Liphardt, J., Dumont, S., Smith, S.B., Tinoco, I., Bustamante, C.: Science 296(5574), 1832 (2002)
102. Go, N.: Annu. Rev. Biophys. Bioeng. 12, 183 (1983)
103. Cho, S.S., Pincus, D.L., Thirumalai, D.: Proc. Natl. Acad. Sci. USA 106, 17349 (2009)
104. Biyun, S., Cho, S.S., Thirumalai, D.: J. Am. Chem. Soc. 133, 20634 (2011)
105. Malo, J., Mitchell, J.C., Venien-Bryan, C., Harris, J.R., Wille, H., Sherratt, D.J., Turberfield, A.J.: Angew. Chem. Int. Ed. 44(20), 3057 (2005)
106. Goodman, R.P., Schaap, I.A.T., Tardin, C.F., Erben, C.M., Berry, R.M., Schmidt, C.F., Turberfield, A.J.: Science 310(5754), 1661 (2005)
107. Seeman, N.C.: Nature 421, 427 (2003)
108. Rothemund, P.: Nature 440, 297 (2006)
109. Yurke, B., Turberfield, A.J., Mills Jr., A.P., Simmel, F.C., Neumann, J.L.: Nature 406, 605 (2000)
110. Green, S.J., Bath, J., Turberfield, A.J.: Phys. Rev. Lett. 101, 238101 (2008)
111. Bath, J., Green, S.J., Allen, K.E., Turberfield, A.J.: Small 5(13), 1513 (2009)
112. Omabegho, T., Sha, R., Seeman, N.C.: Science 324(5923), 67 (2009)
113. Douglas, S.M., Marblestone, A.H., Teerapittayanon, S., Vazquez, A., Church, G.M., Shih, W.M.: Nucl. Acids Res. 37(15), 5001 (2009)
114. Whitelam, S., Feng, E.H., Hagan, M.F., Geissler, P.L.: Soft Matter 5, 1251 (2009)
115. Romano, F., Hudson, A., Doye, J.P.K., Ouldridge, T.E., Louis, A.A.: J. Chem. Phys. 136(21), 215102 (2012)
116. Ouldridge, T.E., Johnston, I.G., Louis, A.A., Doye, J.P.K.: J. Chem. Phys. 130(6), 065101 (2009)
117. Drukker, K., Wu, G., Schatz, G.C.: J. Chem. Phys. 114(1), 579 (2001)
118. Ramachandran, A., Guo, Q., Iqbal, S.M., Liu, Y.: J. Phys. Chem. B 115(19), 6138 (2011)
119. Zou, J., Liang, W., Zhang, S.: Int. J. Numer. Meth. Eng. 0600661, 968 (December 2009, 2010)
120. Lankas, F., Lavery, R., Maddocks, J.H.: Structure 14, 1527 (2006)
121. Harris, S.A., Laughton, C.A., Liverpool, T.B.: Nucl. Acids Res. 36, 21 (2008)
122. Swendsen, R.H.: Phys. Rev. Lett. 42, 859 (1979)
123. Lyubartsev, A.P., Laaksonen, A.: Phys. Rev. E 52, 3730 (1995)
124. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E., Laughton, C.A., Orozco, M.: Biophys. J. 92(11), 3817 (2007)
125. Mazur, A.K.: Biophys. J. 91, 4507 (2006)
126. Galas, D.J., Schmitz, A.: Nucleic Acids Res. 5, 3157 (1978)
127. Tullius, T.D.: Nature 332, 663 (1988)
128. Merino, E.J., Wilkinson, K.A., Coughlan, J.L., Weeks, K.M.: J. Am. Chem. Soc. 127, 4223 (2005)
129. Russell, R., Millett, I.S., Doniach, S., Herschlag, D.: Nat. Struct. Biol. 7, 367 (2000)
130. Jonikas, M.A., Radmer, R.J., Altman, R.B.: Bioinformatics 25(24), 3259 (2009)
131. Parisien, M., Cruz, J.A., Westhof, F., Major, F.: RNA 15(10), 1875 (2009)

132. Tan, R.K.Z., Harvey, S.C.: J. Mol. Biol. **205**, 573 (1989)
133. Wimberly, B.T., Bodersen, D.E., Clemons, W.M., et al.: Nature 407, 327 (2000)
134. Stagg, S.M., Mears, J.A., Harvey, S.C.: J. Mol. Biol. 328(1), 49 (2003)
135. Voltz, K., Trylska, J., Calimet, N., Smith, J.C., Langowski, J.: Biophys. J. 102(4), 849 (2012)
136. Niewieczerzał, S., Cieplak, M.: J. Phys.: Condens. Matter 21(47), 474221 (2009)
137. Górecki, A., Szypowski, M., Długosz, M., Trylska, J.: J. Comput. Chem. 30(14), 2364 (2009)
138. Ouldridge, T.E., Louis, A.A., Doye, J.P.K.: J. Phys.: Condens. Matter 22(10), 104102 (2010)