DNA Duplex Formation with a Coarse-Grained Model

Maciej Maciejczyk,[†,‡] Aleksandar Spasic,[†,§] Adam Liwo,[†,∥] and Harold A. Scheraga*,[†]

[†]Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14850, United States

[‡]Department of Physics and Biophysics, Faculty of Food Sciences, University of Warmia and Mazury, 11-041 Olsztyn, Poland

[§]Department of Biochemistry and Biophysics, University of Rochester Medical Center, Rochester, New York 14642, United States

[∥]Faculty of Chemistry, University of Gdańsk, 80-308 Gdańsk, Poland

S *Supporting Information*

ABSTRACT: A middle-resolution coarse-grained model of DNA is proposed. The DNA chain is built of spherical and planar rigid bodies connected by elastic virtual bonds. The bonded part of the potential energy function is fit to potentials of mean force of model systems. The rigid bodies are sets of neutral, charged, and dipolar beads. Electrostatic and van der Waals interactions are parametrized by our recently developed procedure [Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H.A. *J. Comp. Chem.* **2010**, *31*, 1644]. Interactions with the solvent and an ionic cloud are approximated by a multipole−multipole Debye−Hückel model. A very efficient *R*-RATTLE algorithm, for integrating the movement of rigid bodies, is implemented. It is the first coarse-grained model, in which both bonded and nonbonded interactions were parametrized ab initio and which folds stable double helices from separated complementary strands, with the final conformation close to the geometry of experimentally determined structures.
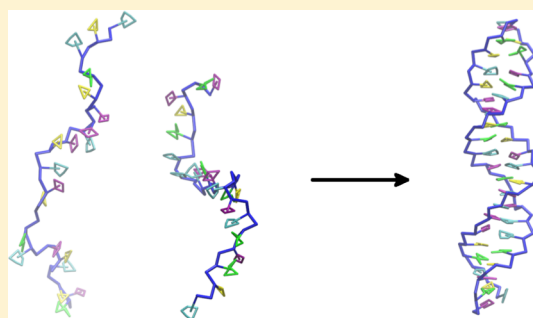
INTRODUCTION

DNA is a blueprint of life. Since Watson and Crick proposed its three-dimensional structure,[1] much effort has been spent to understand its structure and dynamics. Formation of the Watson−Crick double-helix from separated strands is a biologically very important process observed during replication, translation, or DNA repair. It has been the subject of many experimental biochemical and biophysical studies. Also, recently, the rapidly growing field of computer simulations of biological systems has developed models and tools that can provide some insight into this important process.

All-atom molecular dynamics (MD) methods offer the deepest insight into the time evolution of DNA at the atomic level. The CHARMM[2] and AMBER[3] simulation packages have been used to study the duplex dynamics of DNA, mainly in the vicinity of equilibrium. The highlights of near-equilibrium all-atom DNA molecular dynamics simulations include the systematic study of nearest-neighbor effects on various short DNA sequences performed by the Ascona B-DNA Consortium,[4] and the microsecond simulations of the Drew-Dickerson dodecamer[5] or the κB DNA element.[6] Although all-atom MD simulations have proven to be very useful for studying near-equilibrium dynamics, their very high computational cost, related to the large number of degrees of freedom and high-frequency oscillations, forcing a short time-step of integration, currently render them unsuitable for simulating long time-scale nonequilibrium processes, such as formation of duplex DNA from separate strands. Only recently, those enhanced sampling methods, such as replica exchange,[7]

facilitated simulations of formation of very short (4 base pair (bp)) duplexes of DNA.[8] The combination of metadynamics[9] and replica exchange, called BE-META method,[10] was applied to slightly larger system, DNA hexamer,[11] but the more-important biological process of formation of long helices remains prohibitively expensive for all-atom MD simulations and has recently become a target for coarse-graining methodology.

Coarse-grained models can be divided into three categories: statistical models, continuum models, and reduced interaction-center models. In the first category, the early work of Zimm and Rice[12] and Poland and Scheraga[13] neglected all structural and dynamical information and used only the free-energy gain per base-pair formation for computation of the partition function, which, in turn, was transformed into a thermodynamic picture of the process of helix formation. The continuum models approximate dsDNA as a continuum elastic rod and by design cannot be applied to the DNA formation process.[14−16] The last category covers all models in which groups of atoms are replaced by a reduced number of interaction centers connected by elastic and/or rigid virtual bonds.[17−45] A properly designed coarse-graining procedure should reduce the number of interaction centers, leading to a significant speedup of computations, but simultaneously, the simplified potential energy function should include the most important interactions that determine the behavior of the real system. There are two
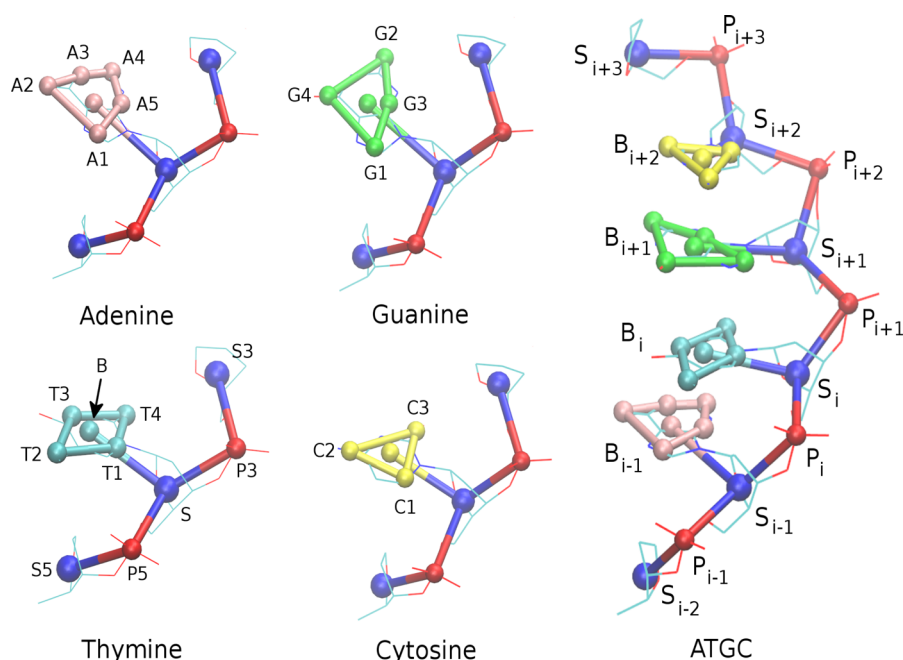
Figure 1. Coarse-grained representation of fragments of the DNA chain superimposed on the atomic representation. Four basic building blocks A, T, G, C attached to small fragments of backbones are shown. Charged beads, which replace phosphate groups, are marked red; neutral beads, which replace deoxyriboses, are marked dark blue. Dipolar beads have different colors for different bases: A, pink; G, green; T, cyan; and C, yellow. Dipolar beads symbols are defined for all building blocks. The internal degrees of freedom used for the definition of the bonded part of the potential energy function (eq 2) are defined in Tables 1−4. For clarity, symbols of beads used for definition of bonded potentials of the backbone are shown for thymine only. The longer fragment of coarse-grained strand of DNA with ATGC sequence overlapped on the atomic model is shown in the right panel. Sequential numbers of consecutive building blocks are assigned.

common approaches to the problem of parametrization of a coarse-grained model.[37] In the "top-down" parametrization method, selected parameters are adjusted to reproduce large-scale properties of the investigated systems. In the "bottom-up" parametrization method, interactions are directly determined from either all-atom molecular dynamics simulations or quantum chemistry computations applied to some model systems.

In the past decade, many coarse-grained models of DNA have been developed, but only several of them managed to form double-stranded DNA from separated complementary chains. A three-bead-per-nucleotide model of de Pablo and co-workers[27] was successfully used to study DNA melting[30] and the renaturation processes.[31] The model was designed around the potential energy function, which effectively depends on the single parameter ϵ, which was optimized to reproduce experimental melting curves. The base−base interactions were approximated by a Gō-like potential.[46]

The model of Ouldridge et al.[37] successfully addresses the phenomena of single-strand stacking, duplex hybridization, and DNA hairpin formation processes. The model was also applied to study DNA Holliday junctions self-assembly[32] and DNA nanotweezers.[36] Although both models successfully fold double-stranded DNA from separate complementary chains, they rely on base−base nonbonded potentials, which either distinguish between native and non-native pairs[27] or nearest-neighbors and all remaining pairs.[32] Such an approach should speed up the in silico folding process, because many "unwanted" free energy local minima are eliminated, but in the real physical system, the nonbonded potential energy of two interacting bases does not depend on their sequential

position in the polynucleotide chain but rather depends on their relative position and orientation in space.

The third model named NARES-2P, which does not incorporate any Gō-like potentials and successfully folds double-stranded DNA from separate strands was recently published by our group.[43] NARES-2P model was developed in parallel to the rigid-body model presented in this paper. NARES-2P is a lower-resolution model with two interaction centers per nucleotide (located on phosphate groups and bases), and it seems to be a minimal physics-based model capable of folding of a double-helix from separate strands. The rigid-body model presented here has more interaction centers and more internal degrees of freedom than NARES-2P, and consequently, it is computationally more demanding. Although the conformational space of rigid-body model is larger than in case of NARES-2P, it is still capable of locating double-helical structures in simulated annealing process started from separate strands. The advantage of the dipolar-bead model over recently published NARES-2P is its higher resolution.

Another model, which is capable of folding dsDNA from separate strands, was published recently by Cragnolini et al.[45] The degree of coarse-graining (6−7 interaction sites per nucleotide) is similar to that of our dipolar-bead model. However, as opposed to our dipolar-bead model, which places more emphasis on the representation of nucleic-acid bases, the model of Cragnolini et al. elaborates on the nucleotide backbone, which is represented by 5 sites per nucleotide than on the nucleic-acid bases, which are represented by 1 or 2 interaction sites, depending on the kind of base; it also utilizes a classical bonded potential and a many-body nonbonded potential responsible for correct Watson−Crick hydrogen-bonding and was parametrized in a "top-down" fashion. This

model does not incorporate any Gō-like potentials and does not treat the interactions between nearest-neighbor nucleotides differently than those between the other pairs of nucleotides.

In this paper, we present a physics-based, middle-resolution model of DNA, which is capable of folding the double-helical structure from separate complementary strands. The model was parametrized in the "bottom-up" fashion, although some adjustment of parameters was necessary to obtain correct balance of key interactions. The non-bonded pairwise potentials do not depend on the sequential position of interacting bases in the polynucleotide chain because the native and non-native, nearest-neighbors, and all other pairs rely on the same potential energy function, which includes a Lennard-Jones and an electrostatic term. The full, physics-based (PB model) version of model keeps stable canonical B DNA double helices in a long room-temperature simulations and folds short DNA molecules. The reduced version of the model, in which base−base intrastrand interactions are limited to nearest neighbors (NN model), successfully folds canonical B DNA from separate complementary strands for all tested systems.

■ METHODS

The Model. The coarse-grained model of DNA is built of three types of particles:

- neutral bead, uncharged Lennard-Jones sphere,
- charged bead, Lennard-Jones sphere with an electric charge located in its center,
- dipolar bead, Lennard-Jones sphere with an electric dipole located in its center.

The DNA chain is built of six types of chemical units: phosphate group (P), deoxyribose (S), adenine (A), thymine (T), guanine (G), and cytosine (C). In the coarse-grained approximation, atoms that constitute each unit are replaced by the three types of beads defined above. The deoxyribose ring is replaced by one neutral bead, the phosphate group is replaced by one charged bead and each base is replaced by a set of dipolar beads. The distances between the dipolar beads of each base are fixed.[47] A schematic representation of the coarse-grained units overlapped on their atomic representation is shown in Figure 1. The parametrization of the bead model is described in the Parametrization section.

Each chemical unit in the coarse-grained approximation constitutes a rigid body. The origin of the local coordinate frame of each rigid body is located at its center of mass, and its axes are aligned with the principal axes of the moment of inertia tensor. The position and orientation of each rigid body with respect to the global coordinate frame is described by a vector−tensor pair $(\mathbf{R},\mathbf{Q})$, where $\mathbf{R}$ is the position of the center of mass and $\mathbf{Q}$ is the rotation matrix. In the model, two types of rigid bodies can be distinguished—spherical (P and S) and planar (A,T,G,C). The spherical rigid bodies interact with other particles only by central forces (see "Potential Energy Function"), and their orientation is irrelevant; therefore, a constant rotation matrix $\mathbf{Q}_{sphere}$ = const is assigned to them. The orientation of each planar rigid body is described by a 3 × 2 rotation matrix $\mathbf{Q}_{planar}$ and varies as the rigid body changes its orientation during a simulation run. The geometry of the chain is defined in vector−tensor space $(\mathbf{R}_i,\mathbf{Q}_i)$ where $i = 1,...,N$ and N is the number of rigid bodies. The relations between global coordinates $(\mathbf{R}_i, \mathbf{Q}_i)$ and distances between various beads, used for definitions of nonbonded interactions, are shown in Figure 2.
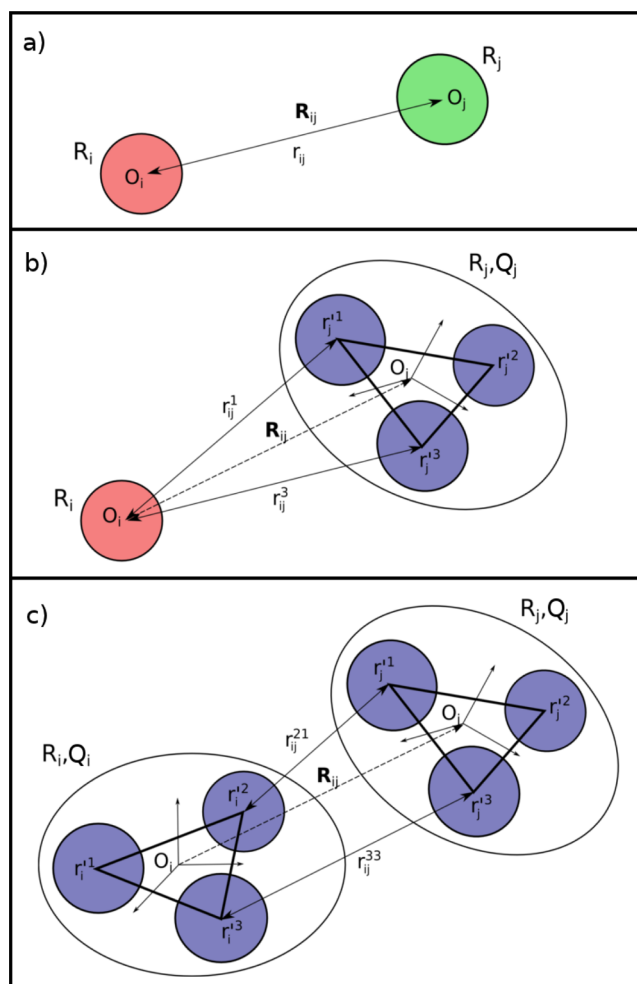

Figure 2. Relations between global coordinates $(\mathbf{R},\mathbf{Q})$ and distances between beads for interactions of (a) two spherical rigid-bodies, where there is only one distance $r_{ij} = |\mathbf{R}_{ij}|$; (b) spherical and planar rigid body, where distances between interacting beads are functions of global coordinates given by eq 9; and (c) two planar rigid-bodies, where distances between interacting beads are given by eq 13.

Rigid bodies are connected by elastic virtual bonds and form a nucleic acid chain as shown in Figure 1. Three rigid bodies connected by two virtual bonds form a virtual-bond angle and four rigid bodies connected by three virtual bonds form a virtual-bond dihedral angle. All virtual bonds, virtual-bond angles, and virtual-bond dihedral angles are listed in Tables 1, 2, and 3, respectively.

Equations of Motion. The algorithm of Leimkuhler and Reich,[48] which is based on the RATTLE[49] constraint method applied to rotation matrices, was used for the description of rotation of rigid bodies. A slightly different version of this

Table 1. Definitions of Virtual Bonds and Parameters of the Bonded Part of the Potential Energy Function[a]

virtual bond	k_d (kcal/(mol Å²))	d_0 (Å)
P5−S	12.45	3.86
P3−S	25.85	3.58
S−B(Pur)	26.55	4.82
S−B(Pyr)	54.23	4.20

[a]For definition of symbols see Figure 1.

Table 2. Definitions of the Virtual-Bond Angles and the Parameters of the Bonded Part of the Potential Energy Function[a]

virtual-bond angle	k_α (kcal/(mol·rad²))	α_0 (rad)
P5–S–P3	10.69	2.11
S5–P5–S	7.59	1.66
P5–S–B(Pur)	7.02	1.83
P5–S–B(Pyr)	8.68	1.62
P3–S–B(Pur)	10.47	1.90
P3–S–B(Pyr)	13.18	2.00
S–B–X(Ade)	32.87	2.71
S–B–Y(Ade)	32.87	1.97
S–B–X(Thy)	42.82	1.10
S–B–Y(Thy)	42.82	2.64
S–B–X(Gua)	31.61	2.22
S–B–Y(Gua)	31.61	0.67
S–B–X(Cyt)	43.95	2.31
S–B–Y(Cyt)	43.95	2.38

[a]S–B–X and S–B–Y are virtual-bond angles between the S–B virtual bond and the X and Y axes of the moment of inertia tensor of a rigid body. (cf. Figure 1)

algorithm, based on the SHAKE[50] (rather than the RATTLE) constraint method, showed superior stability[51] compared to the other commonly used quaternion-based integration scheme. However, this algorithm (based on SHAKE) was not implemented in our work.

The time-evolution of the system is defined in the vector-tensor phase space $(\mathbf{R}_i, \mathbf{p}_i, \mathbf{Q}_i, \mathbf{P}_i)$, where $\mathbf{p}_i$ and $\mathbf{P}_i$ are the linear momentum and generalized angular momentum of the i-th particle, respectively. For a planar rigid body, the rotation matrix is a 3×2 tensor, subject to the orthogonality condition $\mathbf{Q}^T \cdot \mathbf{Q} = \mathbf{I}_2$, where $\mathbf{I}_2$ is a 2×2 unit diagonal matrix.

The leapfrog (Verlet) algorithm was used for propagation of translations, and the *R*-RATTLE (rotational RATTLE) discretization scheme was applied to the rigid-body rotations.[48]

Potential Energy Function. As in our UNRES model of polypeptide chains developed earlier,[52] the effective energy function of our nucleic-acid model stems from the potential of mean force (PMF) of polynucleotide systems immersed in water. The PMF is then to be expanded into a cluster-cumulant series, which, in general, gives rise to the neo-classical local and two-body and nonclassical multibody terms.[52] However, because we keep more explicit degrees of freedom than in the UNRES model, the multibody terms do not seem to be necessary. In particular, in contrast to the UNRES and NARES-

2P models, we do not average over any degrees of freedom of the nucleic-acid bases, the interactions between which are highly directional. In the UNRES model, we did average over the angles of rotation (λ) of the peptide groups about the $C^\alpha ... C^\alpha$ axes; hence, the multibody terms in the UNRES force field are important.[52] In our highly reduced NARES-2P model of nucleic acids,[43] averaging is carried out over the rotation of nucleic bases about their long axes.

The potential energy function of the system is a sum of the bonded and nonbonded parts:

$$U(\mathbf{R}, \mathbf{Q}) = U_{bonded} + U_{nonbond} \qquad (1)$$

Although both the bonded and nonbonded parts depend on the Cartesian-rigid-body degrees of freedom $(\mathbf{R}, \mathbf{Q})$, it is more convenient to define the former as a function of the internal degrees of freedom defined in Tables 1–4 (the internal degrees

Table 4. Definitions of the Virtual-Bond Improper Dihedral Angles and the Parameters of the Bonded Part of the Potential Energy Function[a]

virtual-bond improper dihedral angle	k_γ (kcal/(mol·rad²))	γ_0 (rad)
P5–P3–S–B(Pur)	7.74	2.10
P5–P3–S–B(Pyr)	7.74	1.90

[a]cf. Figure 1.

of freedom are functions of the $(\mathbf{R}, \mathbf{Q})$ vector). The bonded part has a commonly used form

$$U_{bonded} = U_{bond} + U_{bondangle} + U_{dihedralangle} + U_{improper} \qquad (2)$$

where:

$$U_{bond} = \sum_i^{N_{bonds}} k_d(d_i - d_{i0})^2 \qquad (3a)$$

$$U_{bondangle} = \sum_i^{N_{angles}} k_\alpha(\alpha_i - \alpha_{i0})^2 \qquad (3b)$$

$$U_{dihedralangle} = \sum_i^{N_{dihedrals}} \sum_{n=0}^{4} C_{i,n}\cos(n\theta_i) + D_{i,n}\sin(n\theta_i) \qquad (3c)$$

$$U_{improper} = \sum_i^{N_{improper}} k_\gamma(\gamma_i - \gamma_{i0})^2 \qquad (3d)$$

Table 3. Definitions of the Virtual-Bond Dihedral-Angle Parameters of the Bonded Part of the Potential Energy Function[a,b]

virtual-bond dihedral angle	C_0	C_1	C_2	C_3	C_4	D_1	D_2	D_3	D_4
P5–S–P3–S3	0.553	−0.340	0.453	0.001	−0.044	0.102	−0.240	0.090	−0.145
S5–P5–S–P3	0.985	0.422	−0.485	−0.002	−0.053	−0.230	−0.343	−0.084	−0.057
S5–P5–S–B (Pur)	1.168	−0.772	−0.310	−0.008	0.023	0.530	−0.380	0.150	−0.245
S5–P5–S–B (Pyr)	1.246	−0.403	−0.212	0.095	0.082	−0.092	−0.661	−0.038	−0.067
S3–P3–S–B (Pur)	0.478	0.098	0.126	−0.043	−0.028	0.251	0.202	−0.045	−0.033
S3–P3–S-B (Pyr)	0.706	−0.069	0.316	−0.151	−0.125	−0.201	−0.085	0.043	−0.049
P5–S–B–X (Ade)	3.500	3.202				1.412			
P5–S–B–X (Thy)	3.500	3.499				0.731			
P5–S–B–X (Gua)	3.500	−3.497				0.650			
P5–S–B–X (Cyt)	3.500	−3.497				0.153			

[a]P5–S–B–X virtual-bond dihedral angle describes rotation of the base around the S–B pseudoglycosydic bond. [b]cf. Figure 1. All parameters in kcal/mol.

Both bond-stretching and angle-bending terms are approximated by harmonic potentials. k_d and d_{i0} are harmonic force constants and equilibrium distances for bond stretching interactions; k_α and α_{i0} are harmonic force constants and equilibrium angles for the bond-angle bending potential. The general form of the dihedral-angle torsion potentials is a fourth-order Fourier series, although such a long expansion was used only for the backbone dihedral angles (see details in the Parametrization section). $C_{i,n}$ and $D_{i,n}$ are fourth-order Fourier expansion coefficients. The torsional terms are the only higher (second[52]) order terms that are present in the energy function; however, they still belong to the neo-classical term set. The harmonic $U_{improper}$ potential was introduced to maintain the correct position of the center of mass of the base with respect to the sugar ring and surrounding phosphate groups. k_γ and γ_{i0} are harmonic force constant and equilibrium values of the improper dihedral angles. For a detailed geometric definition of the internal degrees of freedom related to the bonded potentials see Figure 1 and Tables 1−4. The force constants and equilibrium positions were determined by fitting the analytical expressions to the potentials of mean force for model systems. The procedure is described in the Parametrization section.

The nonbonded energy is a sum of van der Waals and electrostatic energies and is given by

$$U_{nonbond} = \sum_i \sum_{j<i} U_{P_iP_j} + \sum_i \sum_{j<i} U_{S_iS_j} + \sum_i \sum_{j<i} U_{B_iB_j}$$
$$+ \sum_i \sum_j U_{P_iS_j} + \sum_i \sum_j U_{S_iB_j} + \sum_i \sum_j U_{P_iB_j} \tag{4}$$

where the P, B, and S symbols denote the phosphate group, base, and sugar ring, respectively. The sugar−sugar and sugar−phosphate energy

$$U_{S_iS_j} = U_{LJ}(r_{ij}) \tag{5a}$$

$$U_{P_iS_j} = U_{LJ}(r_{ij}) \tag{5b}$$

include only the Lennard-Jones term and depend only on the distance r_{ij} between two beads. The Lennard-Jones potential has a commonly used form

$$U_{LJ}(r) = \frac{A_{LJ}}{r^{12}} - \frac{B_{LJ}}{r^6} \tag{6}$$

where, A_{LJ} and B_{LJ} are Lennard-Jones coefficients. The phosphate−phosphate energy additionally includes the charge−charge repulsion term

$$U_{P_iP_j} = U_{q-q}(r_{ij}) + U_{LJ}(r_{ij}) \tag{7}$$

The three remaining energy terms of eq 4 involve interactions with the planar rigid-bodies (bases). The sugar−base energy is a sum of excluded volume interactions between a neutral bead and the dipolar beads. This kind of all-repulsive term is also used in the UNRES force field to account for excluded-volume interactions between the peptide groups and side chains[53]

$$U_{S_iB_j} = \sum_k^{N_j} U_{EX}(r_{ij}^k) \tag{8}$$

where N_j is the number of dipolar beads of the j-th base and r_{ij}^k is the distance between the neutral bead of the i-th sugar and

the k-th dipolar bead of the j-th base. It should be noted that r_{ij}^k distances are functions of the distance $\mathbf{R}_{ij}$ between the sugar and the base and the orientation $\mathbf{Q}_j$ of the base (see Figure 2).

$$\mathbf{r}_{ij}^k = \mathbf{R}_{ij} + \mathbf{Q}_j \mathbf{r}'^k_j \tag{9}$$

where $\mathbf{r}'^k_j$ is the position of the k-th dipolar bead in its rigid-body local coordinate frame. The excluded volume potential has a form

$$U_{EX}(r) = \frac{A_{EX}}{r^{12}} + \frac{B_{EX}}{r^6} \tag{10}$$

where the parameters A_{EX} and B_{EX} depend on the type of interacting beads. The phosphate-base interaction additionally includes the charge-dipole terms

$$U_{P_iB_j} = \sum_k^{N_j} [U_{q-d}(\mathbf{r}_{ij}^k) + U_{EX}(r_{ij}^k)] \tag{11}$$

The final term of eq 4 has the most complicated form and describes the interaction between the dipolar beads of two bases

$$U_{B_iB_j} = \sum_k^{N_i} \sum_l^{N_j} [U_{d-d}(\mathbf{r}_{ij}^{kl}) + U_{LJ}(r_{ij}^{kl})] \tag{12}$$

where N_i and N_j are the number of dipolar beads located on the i-th and j-th bases, respectively. The distances r_{ij}^{kl} between dipolar beads are functions of the relative position $\mathbf{R}_{ij}$ and orientations $\mathbf{Q}_i$, $\mathbf{Q}_j$ of the rigid bodies

$$\mathbf{r}_{ij}^{kl} = \mathbf{R}_{ij} + \mathbf{Q}_i \mathbf{r}'^k_i - \mathbf{Q}_j \mathbf{r}'^l_j \tag{13}$$

where $\mathbf{r}'^{k(l)}_{i(j)}$ is the position of the k-th (l-th) dipolar bead in the local coordinate frame of i-th (j-th) base.

The electrostatic interactions are approximated by a multipole−multipole Debye−Hückel model[54]

$$U_{q-q}(r) = D(r) \frac{q^2}{r} \tag{14a}$$

$$U_{q-d}(\mathbf{r}) = -C(r) q \frac{\mathbf{p}_j \cdot \mathbf{n}}{r^2} A(r) \tag{14b}$$

$$U_{d-d}(\mathbf{r}) = C(r) \frac{A(r)\,(\mathbf{p}_i\mathbf{p}_j) - 3B(r)\,(\mathbf{p}_i\mathbf{n})\,(\mathbf{p}_j\mathbf{n})}{r^3} \tag{14c}$$

where $q = e_q$ is the negative unit charge located on phosphate group, $\mathbf{n} = \mathbf{r}/r$ is a unit vector pointing from the first to the second interaction center, and $\mathbf{p}_{i(j)}$ are electric dipole vectors described in the global coordinate frame:

$$\mathbf{p}_i(j) = \mathbf{Q}_{i(j)} \mathbf{p}'_{i(ij)} \tag{15}$$

where $\mathbf{p}'_{i(j)}$ is an electric dipole vector in the local coordinate frame of the i-th (j-th) base. Four functions

$$A(r) = 1 + \kappa r \tag{16a}$$

$$B(r) = 1 + \kappa r + \frac{1}{3}(\kappa r)^2 \tag{16b}$$

$$C(r) = \frac{1}{4\pi\epsilon_0\epsilon_{int}} \exp(-\kappa r) \tag{16c}$$

$$D(r) = \frac{1}{4\pi\epsilon_0\epsilon(r)} \tag{16d}$$

approximate the screening of the electrostatic interactions by ions and solvent, where κ^{-1} is the Debye screening length. For charge–charge interactions distance-dependent dielectric constant of the following form was applied:[40,55]

$$\varepsilon(r) = \begin{cases} \epsilon_{int} & \text{for} \quad r < r_0 \\ \epsilon_{int}\exp[\alpha(r - r_0)] & \text{for} \quad r_0 < r < r_1 \\ \epsilon_\infty \exp(\kappa r) & \text{for} \quad r > r_1 \end{cases} \tag{17}$$

where $\epsilon_{inf} = 78.0$ is a dielectric constant of bulk water and ϵ_{int} is a dielectric constant of the helix "interior", which was taken to be 8.0. The switching distances r_0 and r_1, which determine boundary between unscreened and screened electrostatic interactions, were set to the values of 4 Å and 13 Å.[40] The α parameter was adjusted to keep continuity of dielectric constant.

Because the interactions between the nearest neighbors, that is, particles separated by less than three virtual bonds, are included in the bonded part of the potential, they were excluded from the nonbonded interactions.

Parametrization. *Nonbonded Interactions.* The nonbonded interactions and solvation effects are very important for DNA folding and its conformational stability and were approximated by the excluded-volume potential (eq 10) for P–B and S–B, the Lennard-Jones potential for the remaining nonbonded interactions (eq 6), and a multipole-multipole potential energy function with Debye–Hückel screening. The optimization procedure developed previously[47] was applied to the extended system, which now includes the phosphate group, deoxyribose, and the four bases. The parameters of the B–B interactions, including the positions $\mathbf{r}'$ of the dipolar beads in the local coordinate frame, the electric dipole vectors $\mathbf{p}'$ in the local coordinate frame, and the Lennard-Jones parameters A_{LJ}, B_{LJ}, were retained from the 3445 model of ref 47. For convenience, they can also be found in Tables A-II to A-V of Supporting Information. The missing parameters for the extended system are the charges located on neutral and charged beads, their positions in the local coordinate frame, and the Lennard-Jones parameters of P–P, P–S, S–S, S–B and P–B interactions. The charge of a neutral bead was, by definition, set to zero, the charge of a charged bead was set to $-|e_q|$, which is equal to the charge of the phosphate group in the DNA backbone. The positions of the charges in both the neutral and charged beads were fixed at the center of mass of the corresponding unit. Consequently, the only parameters, which must be determined by the optimization procedure were the Lennard-Jones A_{LJ} and B_{LJ} for P–P, P–S, S–S, S–B, and P–B interactions. The reference energy grids for missing Lennard-Jones interactions were computed in the same manner as the reference grids in ref 47 and the optimization procedure[47] was applied to the extended system.

Although it was initially planned that the Lennard-Jones potential would be applied to all pairwise interactions in the system, extensive testing of dsDNA dynamics showed that its stability was improved when the Lennard-Jones potential of P–B and S–B interactions was replaced by the excluded-volume one. Therefore, the potential was switched to an excluded volume of the form of eq 10, and the A_{LJ} and B_{LJ} parameters were retained from the fitting procedure, that is, $A_{EX} = A_{LJ}$, B_{EX}

$= B_{LJ}$. The values of the computed Lenard-Jones and the excluded volume parameters, presented in the form of $\sigma = (A/B)^{1/6}$ and $\epsilon = B^2/(4A)$, are collected in Table A-I of Supporting Information.

The Debye screening length κ^{-1} was set to 8 Å, which corresponds to the physiological ionic concentration.[56] In this work, the dielectric constant parameter of the bulk water ϵ_∞ was set to 78.0. The stability of the helix and folding efficiency was tested with several values of ϵ_{int} and various distances r_0 and r_1 of the dielectric constant function (eq 17) (see Conformational Stability in the Results and Discussion section).

Bonded Interactions. The virtual bond, virtual-bond angle, and virtual-bond dihedral angle parameters were derived to reproduce the behavior of model systems in the all-atom representation. The interactions between atoms were modeled by the AMBER ff99bsc0[57−59] force field. The potentials of mean force (PMF) of each internal degree of freedom of the coarse-grained model were calculated by using the umbrella sampling method.[60−62] Analytical functions of the form of eqs 3a−3d were then fitted to the numerical PMF's obtained from the WHAM procedure.

The parameters of all virtual bond, virtual-bond angle, and virtual-bond dihedral angle degrees of freedom were computed by using four model systems, which were obtained from three-nucleotide backbones with methyl end groups. The base of the middle nucleotide was either adenine, cytosine, guanine, or thymine, for a total of four different model systems. One such system with a cytosine base in the middle nucleotide is shown in Figure 3. This is a minimal size molecule for which every
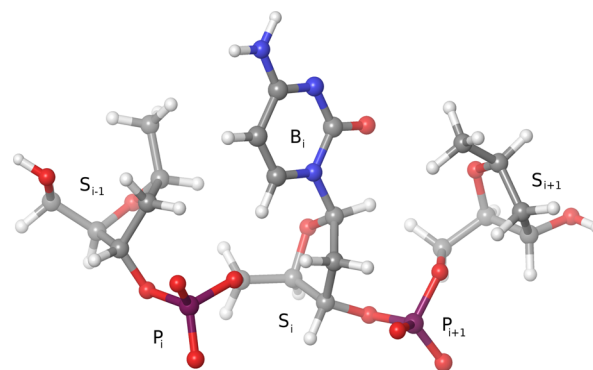

Figure 3. One of four model systems used in deriving the bonded potentials. Model systems for parametrization were created from three-nucleotide chains by removing the bases from the nucleotides at the 5′- and 3′-end and replacing them with methyl groups. The middle nucleotide was left unchanged and can be cytosine (in the figure), thymine, adenine, or guanine.

internal degree of freedom of our coarse-grained model can be defined. Terminal bases were removed to avoid implicitly including electrostatic and van der Waals interactions of bases since our model has the nonbonded parameters derived separately.[47]

Umbrella sampling simulations, described in the Supporting Information, of different model systems produced results that are similar for each internal degree of freedom. Figure A-I in the Supporting Information shows the individual and average PMFs for several virtual bonds, virtual-bond angles, and virtual-bond dihedral angles. The other degrees of freedom from Tables 1−3 show similar trends. Therefore, it was decided to

average the degrees of freedom for purines (model systems with adenine or guanine) and pyrimidines (model systems with cytosine or thymine), except for the degrees of freedom such as the S−P−S virtual-bond angle and the P5−S−P3−S3, S5−P5−S−P3 virtual-bond dihedral angle, which contains two deoxyriboses, and therefore, the base type cannot be uniquely assigned. For these angles, the averaging was done across all four model systems (bases). Some degrees of freedom, which influence orientation of bases with respect to glycosidic bond, were assigned separately for each type of base (see Tables 2 and 3). For each degree of freedom, the distributions of data from each window and for appropriate model systems were combined, the biasing potential was removed, and the PMF along that coordinate was calculated by using the WHAM procedure[60−62] implemented in a program by Dr. A. Grossfield.[63]

The analytical expressions of the form of eqs 3a−3d were then fitted to PMF's. The fitting was done using a trust-region[64] nonlinear least-squares fit algorithm as implemented in MATLAB software. Figure A-II shows the comparison of several PMFs derived from the simulations with the potential calculated using analytical expressions.

Although both the force constants and the equilibrium positions for the virtual bonds and virtual-bond angles were determined by fitting to the PMFs of model systems, it was decided to shift them slightly to match the equilibrium values of the "ideal" B-DNA conformation,[65] in the coarse-grained model. The minima of the virtual-bond dihedral angles were left unmodified. Additionally, harmonic virtual-bond improper P5−P3−S−B with arbitrary force constant and an equilibrium values set to those of the "ideal" B-DNA[65] were incorporated. The rotation of the base around the virtual glycosidic bond (S−B) was restricted by a cosine-shaped potential with the minimum corresponding to the *anti* conformation of the base and the barrier height was preliminary set to 7.0 kcal/mol. The internal degrees of freedom for which the bonded terms were determined and used in the model are listed in Tables 1−4.

We note that a model system and bonded degrees of freedom almost identical to those presented here were used in work by Morriss-Andrews et al.[34] However, the predicted values for the equilibrium values and force constants were somewhat different. There are several reasons for this, first, Morriss-Andrews et al. use statistics from free simulations and CHARM27 force field to calculate the degrees of freedom while this work used umbrella sampling along the degrees of freedom and Amber's ff99bsc0 force field. Most importantly though, the position of the sugar in our model is at the center of mass of sugar ring while they set it at the C1′ atom. This affects not only the equilibrium values but the force constants as well, which makes the direct comparison between bonded degrees of freedom of two models difficult.

Model Systems. The model was tested on the following tree systems, for which the structure was determined experimentally:

- Drew−Dickerson dodecamer, 12 base-pair dsDNA with the sequence 5-CGCGAATTCGC-3′ and its compliment, which has exactly the same sequence (PDB code 1BNA),
- Narayana−Weiss hexadecamer, 16 base-pair dsDNA with the sequence 5′-ACTACAATGTTGCAAT-3′ and its compliment 5′-ATTGCAACATTGTAG-3′ (PDB code 3BSE).

- 21 base-pair dsDNA with the sequence 5′-ACAGCT-TATCATCGATCACGT-3′ and its compliment 5′-ACGTGATCGATGATAAGCTGT-3′ (PDB code 2JYK).

The following 60-nucleotide sequence of ref 34 was used for computation of persistence lengths of single and double stranded DNA:

- HET_{SS} = d(CATCCTCGACAATCGGAACCA-GGAAGCGCCCCGCAACTCTGCCGCGATC-GGTGTTCGCCT)
- HET_{DS} = HET_{SS} + complementary strand

Canonical and Microcanonical Ensemble Simulations. Simulated Annealing. For canonical ensemble simulations, Berendsen's thermal-bath weak-coupling algorithm[66] was implemented. The coupling time was set to 0.1 ps. The initial velocities of the rigid bodies were drawn from the Maxwell distribution.

The stability of the integration algorithm was checked by simulations of the model system in the microcanonical ensemble. The energy of the system was minimized and initial velocities were assigned from the Maxwell distribution; then, the integration of the equations of motion was performed without coupling of the system to the thermal bath.

The ability of the model to fold double-stranded DNA from separated chains was tested with a two-stage simulated-annealing[67,68] procedure. Two complementary strands of each molecule were separated by a random distance and randomly rotated with respect to each other. The separation distance ensured that two strands did not overlap for any random orientation and was different for chains of various lengths. For 1BNA, the separation distance between centers of mass was picked randomly from 40 to 45 Å interval and for 3BSE and 2JYK separation intervals were 50−55 Å and 60−65 Å, respectively. In the first stage, the initial structure was heated up to 600 K (620 for 2JYK) with velocities randomly assigned from Maxwell distribution for 0.1 ns and then cooled to 300 K (320 for 2JYK) with a temperature step ΔT = 50 K. The final temperature was set approximately around 20 K below melting temperature of each molecule. The simulation at each temperature was 0.1 ns long. At the end of each simulation, the presence of contacts between any two bases from separate strands was checked. A contact was assumed to be present when the centers of mass of any two bases from separate chains came within a distances of at most 6.5 Å. If at least one contact was present, the simulation was extended by 1.5 ns for Drew−Dickerson (2.0 ns and 2.5 ns for Narayana−Weiss and 2JYK, respectively) at 300 K in the second stage of the canonical simulation; otherwise, the cycle was terminated and the final structure was excluded from the analysis. The total simulation time of a two-stage simulated annealing cycle was 2.2−3.2 ns (depending on the chain length), and each one was followed by a short energy minimization with Powell's algorithm.[69] The cycle was repeated several thousand times. The number of SA cycles performed for each system is shown in Table 5. Dissociation of chains was prevented by application of a flat-bottom restraint on the distance between their centers of mass with a force constant $k_{CM−CM}$ = 1 kcal/(mol Å²) and a minimum distance of d_0 = 40 Å for 1BNA (50 Å for 3BSE and 60 Å for 2JYK), at which the attractive harmonic force was switched on. The artificial restraint force was switched off when the distance between the centers of mass dropped below 40 Å for 1BNA (50 Å for 3BSE and 60 Å for 2JYK).

Table 5. Statistics of dsDNA Folding for the Nearest-Neighbor (NN) Model[a]

	1BNA		3BSE		2JYK	
trajectories	1352		2428		1937	
long trajectories	670	100%	147	100%	161	100%
0 native contacts	160	24%	63	43%	68	42%
<25% no. contacts	277	41%	54	37%	67	42%
25−50% no. contacts	21	3%	2	1%	2	1%
50−75% no. contacts	54	8%	4	3%	0	0%
>75% no. contacts	158	24%	24	16%	24	15%

[a]The first row shows the total numbers of simulated annealing cycles computed for each molecule. The second row shows the number of trajectories for which the cooling cycle ended up with at least one contact between bases. Statistics of contacts for these trajectories, which were extended by 1.5 ns to 2.5 ns canonical simulation, is shown in the next 5 rows. Numbers of long trajectories with various percentage of native contacts are presented in rows 3−7.

The final structures of the simulated annealing cycle were compared to the experimental reference by computing the all-bead RMSD.

Mechanical Properties of DNA: Persistence Length. The persistence length l_p is a measure of flexibility of a polymer and can be computed from decay of correlation function of unit vectors tangent to helical axis

$$\langle \hat{\mathbf{t}}(s)\,\hat{\mathbf{t}}(0)\rangle = \exp\left(-\frac{s}{l_p}\right) \tag{18}$$

where s is the position along the length of the chain.[70,71]

For ssDNA the tangent vector is defined as a normalized vector between consecutive neutral beads (centers of mass of deoxyriboses). For dsDNA tangent vector is defined as normalized vector connecting midpoints of sugar−sugar distances of complementary bases, separated by 5 base pairs.

Implementation. A new software package for this coarse-grained model, composed of rigid-bodies, was developed. The software is written in C/C++ programming language and currently can perform the following tasks: read the PDB-format structure and transform it into the coarse-grained representation, optimize the geometry of the molecule (by energy minimization), perform microcanonical and canonical ensemble simulations with Berendsen's weak thermal-bath coupling, and perform simulations of heating and simulated annealing cycles. The force and energy calculations were parallelized using OpenMP libraries.[72] The efficiency of the parallelization up to 12 cores is shown in Figure 4. The test was performed on the Drew−Dickerson dodecamer.

■ RESULTS AND DISCUSSION

Numerical Stability, Time Step, and Speedup. The stability of the Verlet-*R*-RATTLE algorithm was tested on the Drew−Dickerson double-helix. Microcanonical simulations of 10 ns length were run with various time-steps: dt = 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 fs. The stability of the total energy is shown in Figure 5. The algorithm maintains good stability up to dt = 22 fs resulting in an order of magnitude speedup compared to the commonly used all-atom time-step of 2 fs (with SHAKE applied on hydrogens), and shows superior stability compared to the quaternion-based rigid-body integration scheme, applied previously[19] to the model of a DNA chain. The drift of the energy over a 20 ns trajectory with dt = 22 fs was only 0.017
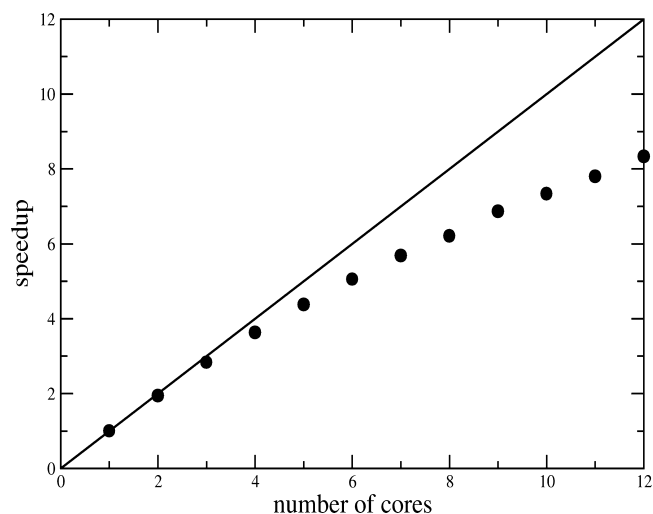

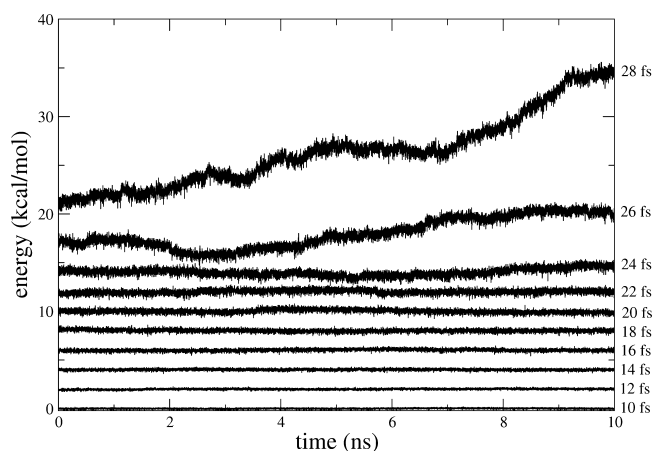**Figure 4.** Efficiency of code parallelization.

Figure 5. Stability of the Verlet-R-RATTLE integration algorithm over 0.1 μs microcanonical simulation of the Drew−Dickerson dodecamer with time-steps ranging from 10 fs to 28 fs.

kcal/mol, compared to ∼10 kcal/mol over 1 ns simulation with the same time-step using a quaternion integration scheme (see Figure 5 in ref 19). The drifts of the energy and the RMS fluctuations for various time-steps are collected in Table 6. All canonical simulations and simulated annealing cycles utilized time-step dt = 10 fs.

The efficiency of the model was compared to that of all-atom simulations performed with the AMBER 11 package. The model system was the Drew−Dickerson dodecamer immersed in a truncated octahedron box filled with 5099 TIP3P water molecules. A 10 Å cutoff and the particle-mesh Ewald summation method were applied for electrostatic interactions. The simulation of 1 ns took 17476 s on 8-cores using Intel Xeon E5645 2.4 GHz CPUs of the computing cluster of the Chemistry Research Computing Facility of Baker Laboratory of Chemistry and Chemical Biology, Cornell University. The same simulation with our coarse-grained model and dt = 20 fs took 235 s with the same CPU, leading to almost 2 orders of magnitude speedup. It should be noted that almost 2 orders of magnitude speedup was achieved without any cutoffs applied to the nonbonded interactions. Implementation of cutoffs, especially to the short-range Lennard-Jones ($U \sim r^{-6}$) and

Table 6. RMS Fluctuations and Drifts of the Total Energies Obtained from 10 ns Simulations of the Drew−Dickerson Dodecamer in the Microcanonical Ensemble[a]

Δt (fs)	10	12	14	16	18	20	22	24	26	28
RMSF (kcal/mol)	0.05	0.06	0.09	0.12	0.15	0.20	0.23	0.39	1.0	1.7
drift (kcal/(mol ns))	0.0002	0.007	0.0011	0.0021	−0.015	−0.015	0.017	0.053	0.371	1.8

[a]RMS fluctuations were calculated for each trajectory with drift subtracted. Drift was calculated as a difference between the averages of the last and the first 1 ns of each trajectory.

the dipole−dipole ($U \sim r^{-3}$) interactions will further increase the speed of computation, especially for long chains.

Conformational Stability. The conformational stability of double-helices was tested with the following procedure. The initial crystal structure (3BNA, 3BSA, and 2JYK) was heated from 100 to 300 K with $\Delta T = 50$ K. The simulation at each temperature was 0.1 ns long, and the production run at 300 K was 0.1 μs long. Snapshots of trajectories were recorded every 1 ps.

The initial simulations of the Drew−Dickerson dodecamer led to a collapse of the helical structure, caused by overestimation of the strength of the Lennard-Jones interactions with phosphates and sugars, which were fitted to reproduce the AMBER potential energy in vacuum. The stability of the system was improved by scaling of Lennard-Jones interaction energy between neutral and charged beads (P−P, S−P, and S−S) by a factor of 0.1.

The stability of a double-helix was also tested with varying r_0 and r_1 parameters of eq 17. Table 7 shows the average number

Table 7. Average Number of Native Contacts Obtained from 0.1 μs Trajectories of Drew−Dickerson Dodecamer for Various Dielectric Constant Switching Values r_0 and r_1 (See Equation 17)[a]

r_0(Å)	r_1(Å)							
	13	14	15	16	17	18	19	20
4	7.2	8.6	8.6	8.2	8.5	9.0	9.1	9.0
6	8.0	6.7	7.8	8.9	9.4	9.0	9.5	8.6
7	7.0	8.6	9.2	8.9	9.7	9.9	9.5	9.9
8	8.3	7.8	8.3	9.3	9.9	10.0	9.6	10.2

[a]The dielectric constant of helix interior was set to 8.0.

of native contacts as a function of distances r_0 and r_1 for Drew-Dickerson dodecamer with $\epsilon_{\text{int}} = 8.0$. The largest average number of native contacts is located in the lower right corner of Table 7, which means that sufficient electrostatic repulsion of

phosphate groups of interacting chains is necessary for stability of double-helix. Smaller switching distances r_0 and r_1 causes instability of the helix termini, which can even destabilize helix interior. On the other hand the reduced (and less physical) model, in which intrastrand base−base interactions were limited only to nearest-neighbors, was tested. The NN model generates very stable double helices even for short switching distances $r_0 = 4$ Å and $r_1 = 13$ Å, as applied in the ref 40. It seems that stronger electrostatic repulsion of phosphate groups is necessary to prevent deformations of helix termini caused by base−base intrastrand interactions. On the other hand, stronger electrostatic repulsion affects folding rate in the fast simulated annealing procedure (see section Folding of dsDNA from Separate Chains).

Table 8 collects mean helical parameters obtained from trajectories for three model systems for full physics-based model and for the simplified nearest-neighbor model. These parameters were calculated from coarse-grained representations of experimental reference structures and the structures obtained in simulations, respectively. A precise determination of all helical parameters would require conversion of the coarse-grained structurs to an all-atom representation; however, the values calculated by using a coarse-grained representation are sufficient for comparison of the simulated structures with the respective experimental structures. For all model systems, the average number of native contacts is larger for the NN model. For shorter molecules, 1BNA and 3BSE, the radius of gyration within NN approximation is close to the value obtained for experimental structure. For the PB model applied to 1BNA and 3BSE the radius of gyration is slightly too short, which is caused by deformation of helix termini. For the longest molecule both models overestimates the radius of gyration. Both models underestimate helix radius and overestimate helical rise for all dsDNAs, although these discrepancies are relatively small. Larger deviations are observed for the size of minor groove, which is significantly overestimated by both models. The major groove size is relatively well preserved.

Table 8. Mean Parameters Describing the Geometry of the Helix During 0.1 μs Simulation and Comparison to the Crystal Structure Geometry[a]

param. (Å)	native contacts	radius of gyration	helix diameter	rise[b]	major groove width	minor groove width
Drew−Dickerson (PB)	7.3(2.0)	12.5(0.4)	18.1(0.9)	3.4(0.6)	15.5(2.2)	15.5(2.0)
Drew−Dickerson (NN)	10.5(1.2)	13.5(0.3)	18.1(0.7)	3.7(0.4)	16.2(2.6)	15.6(1.8)
Drew−Dickerson (Exp)	12.0	13.2	18.9(0.6)	3.5(0.2)	17.5(0.5)	11.1(1.8)
Narayana−Weiss (PB)	11.8(1.9)	16.6(0.5)	18.2(0.8)	3.5(0.5)	15.9(2.3)	15.7(2.1)
Narayana−Weiss (NN)	13.3(1.6)	17.1(0.4)	18.3(0.7)	3.7(0.4)	16.3(2.2)	15.8(2.0)
Narayana−Weiss (Exp)	16.0	17.1	19.2(1.0)	3.4(0.1)	18.2(0.8)	11.6(1.2)
2JYK (PB)	16.3(1.9)	20.7(0.6)	18.3(0.7)	3.5(0.6)	16.2(2.5)	15.7(2.0)
2JYK (NN)	18.0(1.6)	21.4(0.6)	18.3(0.7)	3.7(0.4)	16.6(2.4)	15.8(1.9)
2JYK (Exp)	21.0	19.0	19.3(0.2)	3.5(0.3)	17.7(1.0)	10.8(1.3)

[a]Numbers in parentheses are standard deviations. PB and NN denotes the full physics-based and the nearest-neighbor model, respectively. [b]The helical rise is approximated by the distance between consecutive centaral points of lines connecting centers of mass of Watson−Crick-paired bases.

Persistence Length. Persistence lengths of both HET_{SS} and HET_{DS} were computed from 0.1 μs canonical simulations at 300 K with 145 mM salt concentration for both full PB and simplified NN model. The results are presented in Table 9. For

Table 9. Persistence Lengths of HET60$_{SS}$ and HET60$_{DS}$ for Full PB and Reduced NN Models[a]

persistence lengths (Å)	HET60$_{SS}$	HET60$_{DS}$
full PB model	6.5	151.7
reduced NN model	28.3	158.3

[a]Persistence lengths were obtained by fitting of eq 18 the the correlation functions obtained from canonical ensemble trajectory computed at 300 K with 145 mM salt concentration.

both chains PB model produces smaller persistence lengths than NN model, although the relative difference is much bigger for ssDNA. This behavior is expected, because switching off some of the intrastrand base—base interactions should decrease bending of the chain, especially for ssDNA, in which bases are exposed to solvent and can easily form intrastrand hairpins with complementary bases. Figure 6 presents typical configurations
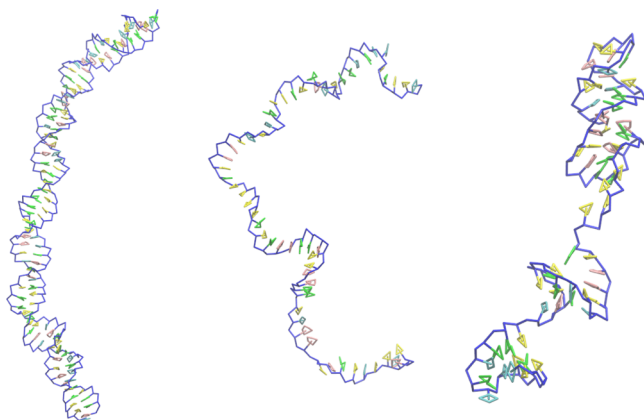
Figure 6. Typical configurations of HET60$_{DS}$ (left panel), HET60$_{SS}$ within nearest-neighbor approximation (middle panel) and HET60$_{SS}$ for full physics-based model (right panel).

of HET60$_{SS}$ and HET60$_{DS}$ molecules taken from $T = 300$ K canonical simulation. The conformation of ssDNA taken from full PB model simulation differs significantly from the one taken from NN approximation. The ssDNA obtained from full PB model bends and forms hairpins, which can act as kinetic traps in the folding process of dsDNA.

The value of persistence length of ssDNA obtained with NN model agrees reasonably well with experimental values.[73] The persistence length of dsDNA appears to be around three times too small compared to experimental values,[74−77] and it is close to the value computed by early version of 3SPN model.[27] Although the agreement with experimental data is not perfect it appears to be quite satisfactory for "bottom-up"-parametrized physics-based coarse-grained model and this discrepancy will be corrected in next versions of the model.

Mismatches. The influence of mismatches was tested on the 30-bp system of 10 G/G pairs flanked by 10 A/T pairs at both 5′ and 3′-end. During the 0.1 μs canonical simulation at 300 K, opening of mismatched G/G pairs was not observed. Although interaction of complementary bases is different than that between mismatched bases, the model does not seem to be

very specific with respect to base—base interactions. This problem may be result of exaggeration of base—base Lennard-Jones interactions with respect to electrostatics and will be addressed in the next version of the model.

Folding of dsDNA from Separate Chains. The simulated annealing (SA) procedure described in the Methods section was applied to some model systems. The folding efficiency of the full PB model was low. Only 3 out of 2100 trajectories led to the final helical structures with more than 10 native base—base contacts for Drew−Dickerson dodecamer. Longer molecules did not fold at all. Much better folding rate was achieved with NN model in which intrastrand kinetic traps (hairpins), resulting from intrastrand long-range base—base interactions, were eliminated. The statistics of the folding process of three model systems within NN approximation are presented in Table 5. After the majority of simulated annealing cooling cycles (first stage), the chains did not come into contact because of electrostatic repulsion of the highly charged chains and the relatively short time of the simulation. The frequency of making at least one contact after a cooling cycle is different for each simulated system and changes from 50% for the dodecamer to 6% for the hexadecamer, and to 8% for the 21-bp system. This phenomenon can be explained by increasing volume of an accessible space for longer chains, which decreases probability of making contact during fast cooling process. The effect is enhanced by increased electrostatic repulsion of the longer negatively charged backbones.

After the second stage of simulated annealing, native-like structures were obtained. The percentage of the native-like structures, based on the number of native contacts, is shown in Table 5. It should be noted that the numbers summarized in Table 5 were obtained from SA calculations which do not reflect the actual amount of the native-like structures that a force-field can produce. The cooling time is always too small to obtain an equilibrated ensemble. The amounts of contacts in Table 5 are, therefore, underestimated.

Nevertheless 24% of extended trajectories led to double-helical structure with more than 10 native contacts for 1BNA molecule. Roughly the same number of structures broke up into separate chains or formed misfolded structures without native contacts. 41% trajectories ended up in misfolded structures with 1−3 native contacts, which were mostly parallel dsDNA's, in which 5′-end of one chain makes nonative contacts with 5′-end of the other chain. Figure 7 shows the potential energy of final structures as a function of all-bead RMSD computed with respect to experimental structure. Two low-energy clusters of structures are clearly visible. The one located in the lower left corner of Figure 7 is a cluster of correctly folded antiparallel dsDNA's and the other one represents misfolded parallel dsDNA's, in which 3′-ends (and 5′-ends) of chains are paired. The important observation that comes from this graph is that lowest energy structure is clearly located in the correctly folded group and its energy is around 10 kcal/mol lower than misfolded parallel dsDNA, although the lowest energy structure is not the structure with lowest RMSD. The mean potential energy of antiparallel dsDNA cluster is around 8 kcal/mol lower than the mean energy of the parallel DNA cluster, which clearly favors correctly folded structures, although this difference for real systems might be larger (experimental structures of the parallel Drew−Dickerson dodecamer were not observed) and further refinement of the electrostatics and/or Lennard-Jones balance of the model is required to increase this gap and improve specificity of base—base interactions.
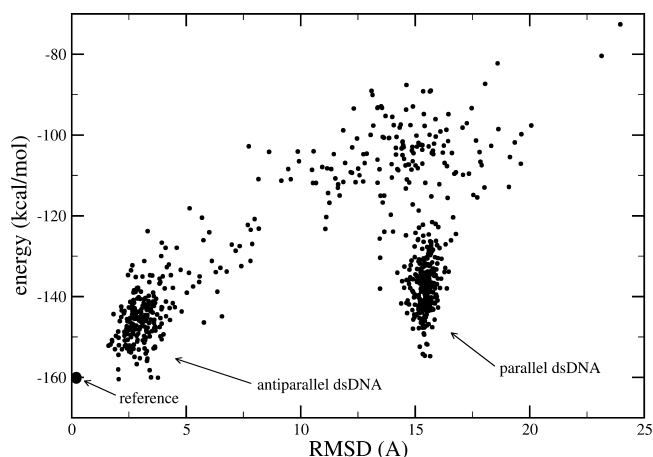
Figure 7. Energy vs RMSD with respect to the experimental structure for the 1BNA dodecamer. Two major clusters, corresponding to antiparallel and parallel dsDNA, and the energy-minimized experimental structure are pointed by arrows.

The reference structure marked by bigger dot in Figure 7 is the energy-minimized experimental structure. The optimization of experimental structure was necessary to remove small steric clashes in original structure caused by imperfectness of the coarse-grained model. The potential energy of original 1BNA molecule is more than 100 kcal/mol higher than that of the optimized one, but the all-bead RMSD between the optimized and original structure is only 0.4 Å. This effect can be expected for the coarse-grained model with force-field parameters developed in the "bottom-up" fashion. The potential energy of the reference structure is almost exactly the same (<1 kcal/mol) as the lowest-energy structure obtained from simulated annealing procedure.

The folding efficiency of longer polymers decreases as the dimensionality of phases spaces increases (see Table 5). 16% of long trajectories of 3BSE molecule led to final structures with more than 75% of native contacts (13−16 contacts). For the 2JYK molecule, the folding rate is almost the same as that for 3BSE and equals 15%. Nevertheless, the energy vs RMSD graphs (Figures 8 and 9) are qualitatively similar to these obtained for 1BNA molecule (Figure 7). In both cases, two clusters of antiparallel and parallel dsDNAs are clearly visible and correct—antiparallel cluster has mean potential energy
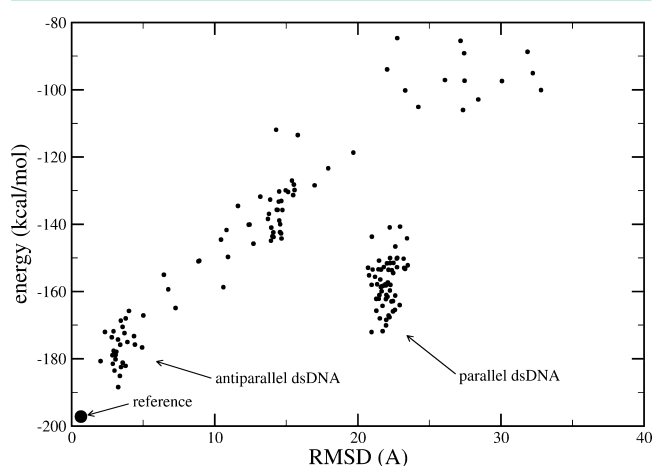

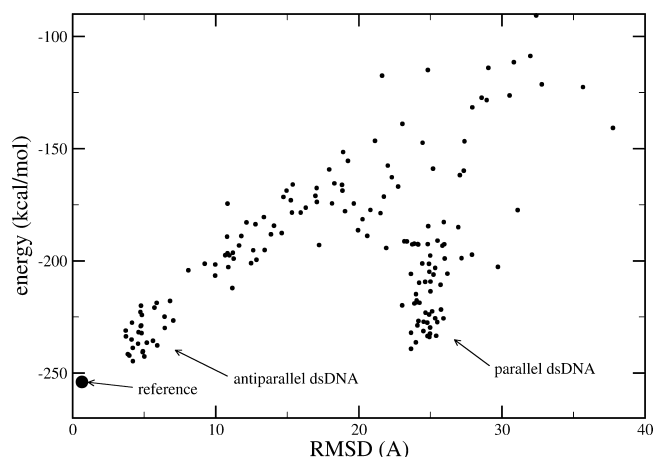**Figure 8.** Same as Figure 7 but for the 3BSE hexadecamer.

Figure 9. Same as Figure 7 but for the 2JYK molecule.

around 20 kcal/mol lower than this of incorrect—parallel dsDNA cluster. These values should effectively discriminate misfolded structures in the real-, long-time folding process. This mean-energy difference increases to 22 kcal/mol for the longest 2JYK molecule. The energy-minimized reference structures, marked by bigger dot in Figures 8 and 9, have RMSD only 0.6 Å away from original structure. In both cases, the reference structures have potential energy around 10 kcal/mol lower than these obtained from simulated annealing procedure. This observation suggests that longer cooling process should lead to improvements in geometry of final structures, as the minimum of potential energy was not reached by any simulated annealing cycle for longer model systems.

Figure 10 presents the three lowest-energy structures of tested molecules superimposed on their experimental structure. The all-bead RMSDs of 1BNA, 3BSE, and 2JYK molecules with respect to experimental references are 2.1 Å, 3.1 Å, and 4.2 Å, respectively, and as expected, they increase with the length of polymers. All structures clearly form double-helices, although some discrepancies of the geometry are visible, especially for the longest molecule, for which overestimation of the size of minor groove is clearly visible.

Analysis of 158 trajectories of 1BNA, which led to structures with 10−12 native contacts revealed two qualitatively different mechanisms of folding.

The representative of the first most common mechanism (93 of 158 trajectories) is shown in Figure 11. Two separated chains (0 ns) were heated to 600 K and evolved into random chains (0.11 ns, T = 550 K). At 0.56 ns (T = 350 K) the 5′-GCGC-end established native contacts with the complementary 3′-CGCG-end, and since then, more native contacts were formed, and the dsDNA molecule was zipped. At 0.68 ns, all native Watson−Crick hydrogen bonds were formed. This type of folding mechanism is referred to as zippering.[31]

A different and less common folding process (65 of 158 trajectories) is shown in Figure 12. It was started from two separated ssDNA chains, which were heated to 600 K and, after 0.3 ns (T = 500 K), evolved into a completely random conformation. First, contacts between the bases of the 5′-ends of the chains were established after 0.49 ns. It should be noted that these are non-native Watson−Crick G/C contacts. After that, the chains started to slide with respect to each other along a Watson−Crick interface. After 0.89 ns, two G/C contacts were estabilished but also A/C and A/G mismatched pairs were formed. A 1.4 ns snapshot shows a molecule with 10 contacts
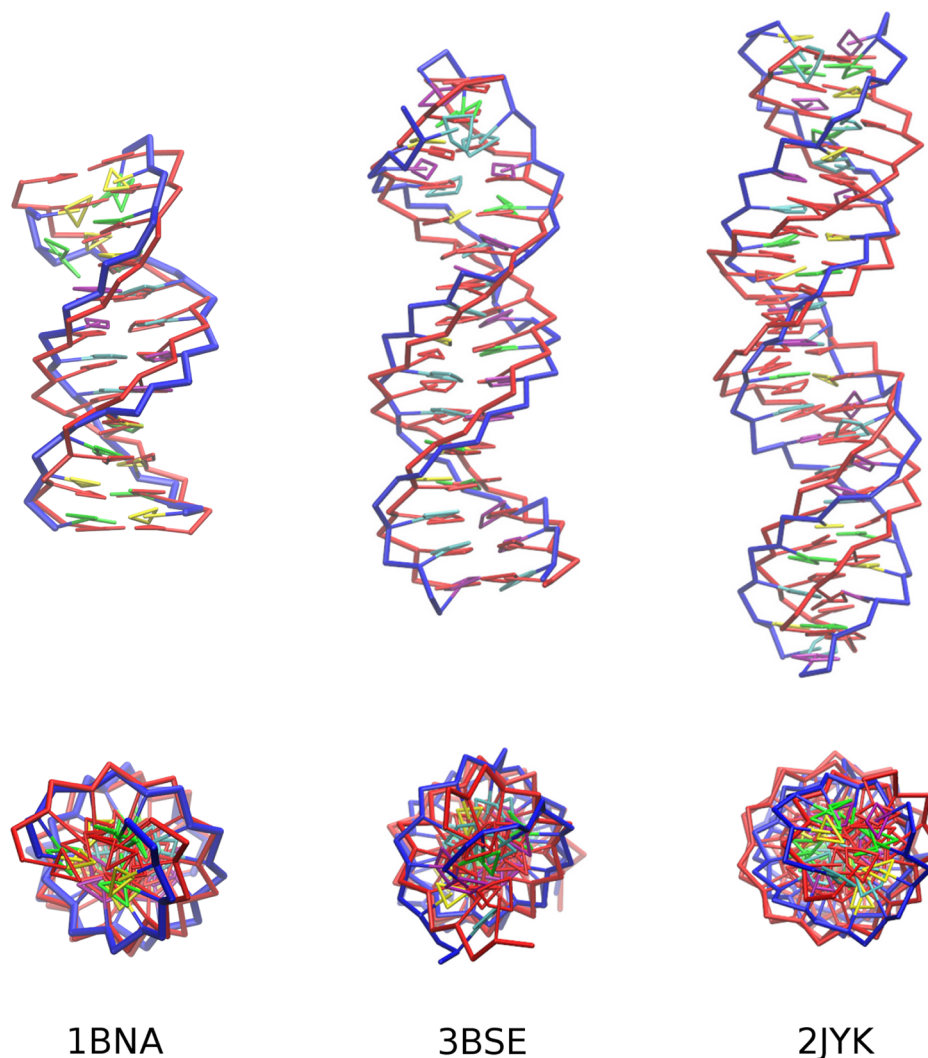
1BNA   3BSE   2JYK

Figure 10. Lowest-energy structures of 1BNA, 3BSE, and 2JYK molecules obtained from extended simulated annealing procedure, superimposed on experimental structures. Their all-bead RMSDs are 2.1 Å, 3.1 Å, and 4.2 Å, respectively. The reference structures are shown in red. The backbone of the computed structures is shown in blue and the bases are A (cyan), T (pink), G (green), C (yellow).

established and with distorted termini. After 1.7 ns a full helix with all native base–base contacts was formed. This type of folding mechanism is referred to as slithering.[31]

As can be expected, the zippering mechanism was also observed for longer molecules, 3BSE and 2JYK. Nevertheless, some trajectories with a slithering folding mechanism were also recorded. At first, this observation is surprising because, as opposed to the 1BNA molecule, the initial mis-alignment of the two strands creates a number of mismatched base–base pairs. Therefore, the fact that the simulated folding of 3BSE and 2JYK proceeded almost as frequently by the zippering and by the slithering mechanism probably resulted from low base–base pairing specificity of the model. The snapshot from helix initiation process for 3BSE molecule is shown in Figure 13a). The 3′-end of a complementary strand made contacts with 3′-end of the leading strand. Besides three nonative Watson–Crick contacts (two A/T and one G/C) and two mismatched contacts (G/T and T/T) were established. Three Watson–Crick pairs and two mismatched pairs form initial structure from which monomers start slithering with respect to each other until final proper double-helix is formed. The helix nucleation occurred between nucleotides of 3′-ends but also the

mechanisms where 3′-end (or 5′-end) of one chain made a contact with midsection of the other chain were observed.

For the longest chain, the slithering mechanism was observed only for two out of 22 folding trajectories. In one case, nucleation started from formation of contacts of 3′-end of one chain with middle section of the other chain (see Figure 13b). In the second slithering trajectory, the 5′-end made initial contact with middle section of the other chain. No trajectory where two 3′-ends (or 5′-ends) made initial contact was recorded.

At this point, the analysis of DNA duplex hybridization is reduced only to qualitative observations presented above. Further development of the model is necessary, especially further reduction of computational cost via implementation of cutoffs, to obtain large set of trajectories, which is required for proper analysis of thermodynamics and kinetics of DNA hybridization process.

Nevertheless, the mechanism of hybridization can be qualitatively compared to hybridization mechanisms proposed by de Pablo[30] and Doye[44] groups. The former group showed two mechanisms also called slithering and zippering, which follows double-helix nucleation process.[30] The former mecha-
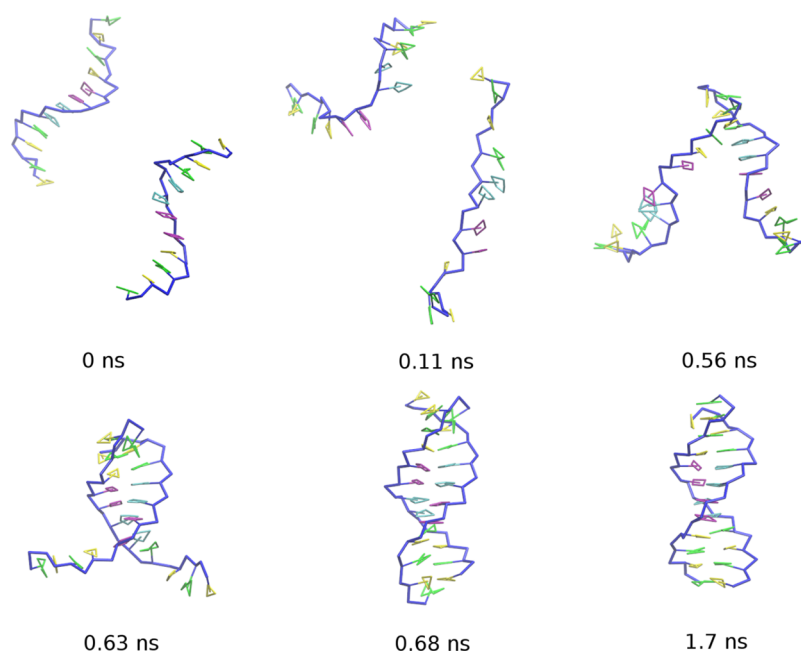
5031

dx.doi.org/10.1021/ct4006689 | *J. Chem. Theory Comput.* 2014, 10, 5020–5035

Figure 11. Snapshots of trajectory of a zippering mechanism of the folding process of the Drew−Dickerson dodecamer.
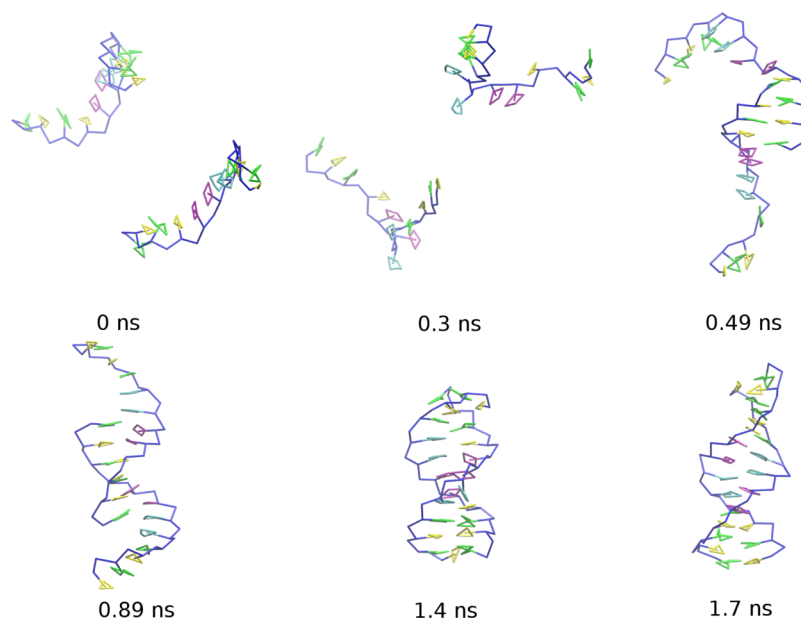

Figure 12. Snapshots of trajectory of a slithering mechanism of the folding process of the Drew−Dickerson dodecamer.

nism is dominant for repetitive sequences and zippering is characteristic of random sequences. Both mechanisms seem to be similar to folding mechanisms obtained with our dipolar bead model, although it should be noticed that sequences of simulated systems are not exactly the same. The Drew−Dickerson dodecamer is symmetric and might have pathways of folding that are similar to slithering mechanism observed for 3SPN model. Our model shows that slithering and zippering mechanisms for 1BNA are almost equally probable. The 3BSE and 2JYK molecules have a sequence that can be described as random, and their dominant folding mechanism is molecular zippering.

The zippering mechanism was also observed for the model of Ouldridge et al.[44] They showed that there is a bias toward initial structures with contacts between ends of the strands, an observation that can be confirmed by our model. Almost all zippering trajectories were initiated by contacts of monomers termini. One of the mechanisms of duplex formation observed for repetitive sequences, so call inchworm mechanism, was not observed in our simulations, although we did not test polymers with exactly the same sequence. The inchworm mechanism may also be related to the specificity of base−base interactions. It should also be noticed that tendency for slithering, instead of inchworm-like behavior, in our model may result from insufficient specificity of base−base interactions. This problem will be addressed in the future development of the model.

Finally, it should be stressed that these results were obtained in unrestrained simulations started from separated chains. This mode of simulations is different from that of many other works in which simulations are started from the experimental
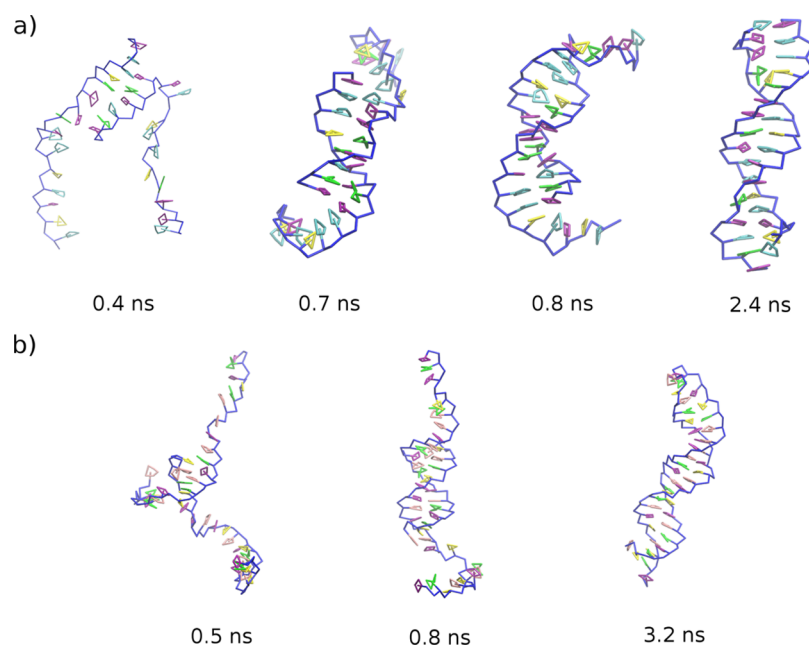
Figure 13. Examples of slithering pathways of DNA hybridization of (a) 3BSE and (b) 2JYK molecules.

structure. Unlike many other force-fields developed for DNA, our force-field, therefore, has ab initio folding capacity.

Influence of Backbone Dihedral Angle Potentials on the Formation of Double-Helix. In our recent paper, it was shown that dihedral angle bonded potentials are not prerequisites for double-helix formation, which is rather driven by base−base dipole interaction.[43] Also, the model developed by Ouldridge et al. does not require backbone dihedral angle potentials to obtain DNA duplex hybridization.[36] A question can, therefore, be asked: Are the backbone potentials a necessary factor for double-helix formation, or is this process solely dependent on the base stacking and pairing? The problem was addressed by running the simulated annealing procedure with P5−S−P3−S3, S5−P5−S−P3, S5−P5−S−B, and S3−P3−S−B dihedral angle potentials switched off. The test performed on the Drew−Dickerson dodecamer showed the same efficiency of double-helix hybridization process (24%). The lowest energy structure shown in Figure 14a is only 3.2 Å away from the experimental structure. Next, improper rotamer-like potentials P5−P3−B−S, P5−S−B, and P3−S−B were additionally switched off. The lowest energy structures have a ladder-like shape, as shown in Figure 14b. It seems that backbone dihedral angle potentials are not really necessary for DNA duplex hybridization. Switching off rotamer-like potentials (P5−S−B, P3−S−B, and P5−P3−S−B) causes formation of ladder-like structures instead of double-helices. This effect might be related to the lack of specificity of base−base interactions, which might be the result of overestimation of base−base van der Waals interactions with respect to dipole−dipole electrostatic interactions. Nevertheless, the dipolar-bead model suggests that only backbone bond-stretching, angle-bending, and rotamer-like potentials are necessary for double stranded DNA hybridization.

■ CONCLUSIONS

A physics-based rigid-body coarse-grained model of DNA is proposed. The model was parametrized in the "bottom-up" fashion. The bonded interactions are fitted to the potentials of
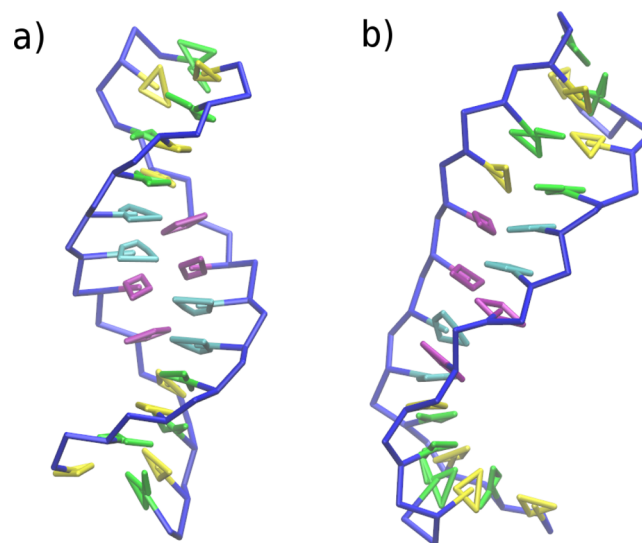

Figure 14. Lowest energy structures obtained from simulated annealing procedure with the NN model with some of bonded potentials switched off. (a) Dihedral angle potentials S5−P5−S−P3, P5−S−P3−S3, S5−P5−S−B, and P3−S3−S−B switched off. (b) Additionally, rotamer-like potentials P5−S−B, P3−S−B, and P5−P3−S−B switched off.

mean force of the model system. The equilibrium virtual-bond lengths and virtual-bond angles were set to the values of the ideal B-DNA double-helix. The nonbonded interactions are approximated by Lennard-Jones, excluded volume and electro-static interactions of charges and dipoles. The solvation and ionic cloud screening is approximated by the Debye−Hückel model. The model does not incorporate any Gō-like potentials, which force the structure toward the experimental one. In the full physics-based model, nonbonded potentials do not distinguish between near-neighbors and any other base pairs. A full physics-based version of the model the nearest-neighbor approximation, in which intrastrand base−base interactions are limited to nearest-neighbors only, was tested. It should be

stressed that in the reduced model all interstrand base−base interactions are present.

The *R*-RATTLE rigid-body integration algorithm, implemented in our software package, demonstrated superior stability compared to a quaternion-based algorithm, with low drift of the total energy even for a time-step as long as 20 fs. The total speedup of computations compared to all-atom simulations is close to 2 orders of magnitude. This result was achieved without application of cutoffs on the nonbonded interactions.

Stable double-helix trajectories were obtained for both PB and NN models after adjustment of several force-field parameters as discussed in the Conformational Stability section. The geometry of two experimental DNA structures is preserved during 0.1 μs simulations at 300 K. The persistence lengths of ssDNA and dsDNA are underestimated by the factor of 2−3 compared to experimental data, discrepancy that seems to be reasonable for physics-based model designed in the "bottom-up" fashion. The other drawback of the model, which seems to have the same origin as discrepancy in persistence lengths, is insufficient specificity of base−base interactions, which affects simulations of dsDNA with mismatches. Opening of mismatched pairs was not observed during 0.1 μs canonical simulations at 300 K.

Nevertheless, the model successfully folds short (12−21 bp) DNAs from separated strands. The full PB model is only able to fold shortest double-stranded DNA dodecamer from separated chains with relatively low efficiency caused by formation of intrastrand hairpins—kinetic traps in the folding process. The reduced NN model removes kinetic traps and greatly improves the folding efficiency. All three tested molecules with lengths varying from 12 to 21 bp were folded from separate strands by NN model, although the efficiency of folding drops with increasing size of the simulated system. The possible explanation is that the cooling process is too fast to sample conformational space effectively and find the global minimum on the complicated multi-minima free energy hyper-surface. Application of a replica-exchange method, which will be implemented in our software package, should help to explain the cause of the low helical populations produced by the simulated annealing procedure.

The model predicts two mechanisms of dsDNA hybridization. The more frequent process called zippering starts from formation of native base−base contacts between 3′ and 5′-ends of two chains and is followed by helix propagation. In the second, less frequent process called slithering, initial non-native Watson−Crick contacts (also mismatched contacts for 3BSE and 2JYK) are established between 5′-ends (or 3′-ends) of two strands; then, chains slide with respect to each other along the Watson−Crick interface until all native contacts are formed. This process was observed for all model systems although it is much more frequent for the dodecamer, which has specific sequence, than for other molecules. The slithering process might also be an artifact caused by insufficient specificity of base−base interactions.

Finally, the influence of backbone dihedral angle and rotamer-like potentials on a double-helix formation process was tested. The backbone dihedral potentials do not seem to be a necessary prerequisite for dsDNA hybridization, but without rotamer-like potentials, ladder-like structures instead of double helices were observed.

The model has some weaknesses, such as low specificity of base−base interactions, underestimation of persistence lengths of both single and double-stranded DNA, and small influence of mismatched pairs on double-helix dynamics. The origin of mentioned problems most probably arises from overestimation of Lennard-Jones base−base interactions, and we will address these problems in the next evaluations of the dipolar bead model.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Umbrella sampling procedure applied to model systems; graphs of selected potentials of mean force and their best fits of analytical functions and all nonbonded parameters of neutral, charged, and dipolar beads. This material is available free of charge via Internet at http://pubs.acs.org

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 607 255 4034. Fax: 607 254 4700. Email: has5@cornell.edu.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737−738.

(2) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(3) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(4) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., III; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J. *Nucleic Acids Res.* **2010**, *38*, 299−313.

(5) Perez, A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2007**, *129*, 14739−14745.

(6) Mura, C.; McCammon, J. A. *Nucleic Acids Res.* **2008**, *36*, 4941−4955.

(7) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(8) Kannan, S.; Zacharias, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10589−10595.

(9) Laio, A.; Parinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.

(10) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.

(11) Piana, S. *J. Phys. Chem. A* **2007**, *111*, 12349−12354.

(12) Zimm, B. H.; Rice, S. A. *Mol. Phys.* **1960**, *3*, 391−407.

(13) Poland, D.; Scheraga, H. A. *Physiol. Chem. Phys.* **1969**, *1*, 389−446.

(14) Marko, J. F.; Siggia, E. D. *Macromolecules* **1994**, *27*, 981−988.

(15) Balaeff, A.; Mahadevan, L.; Schulten, K. *Phys. Rev. Lett.* **1999**, *83*, 4900−4903.

(16) Balaeff, A.; Koudella, C. R.; Mahadevan, L.; Schulten, K. *Philos. Trans. R. Soc., A* **2004**, *362*, 1355−1371.

(17) Vorobjev, Y. N. *Biopolymers* **1990**, *29*, 1503−1518.

(18) Olson, W. K. *Curr. Opin. Struct. Biol.* **1996**, *6*, 242−256.

(19) Rudnicki, W. R.; Bakalarski, G.; Lesyng, B. *J. Biomol. Struct. Dyn.* **2000**, *17*, 1097−1108.

(20) Maciejczyk, M.; Rudnicki, W. R.; Lesyng, B. *J. Biomol. Struct. Dyn.* **2000**, *17*, 1109−1116.

(21) Drukker, K.; Wu, G.; Schatz, G. C. *J. Chem. Phys.* **2001**, *114*, 579−590.

(22) Jendrejack, R. M.; de Pablo, J. J.; Graham, M. D. *J. Chem. Phys.* **2002**, *116*, 7752−7759.

(23) Jendrejack, R. M.; Schwartz, D. C.; Graham, M. D.; de Pablo, J. J. *J. Chem. Phys.* **2003**, *119*, 1165−1173.

(24) Jendrejack, R. M.; Schwartz, D. C.; de Pablo, J. J.; Graham, M. D. *J. Chem. Phys.* **2004**, *120*, 2513−2529.

(25) Chen, Y. L.; Graham, M. D.; de Pablo, J. J. *Macromolecules* **2005**, *38*, 6680−6687.

(26) Tepper, H. L.; Voth, G. A. *J. Chem. Phys.* **2005**, *122*, 124906.

(27) Knotts, T. A., IV; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. *J. Chem. Phys.* **2007**, *126*, 084901.

(28) Voltz, K.; Trylska, J.; Tozzini, V.; Kurkal-Siebert, V.; Langowski, J.; Smith, J. *J. Comput. Chem.* **2008**, *29*, 1429−1439.

(29) Niewieczerzal, S.; Cieplak, M. *J. Phys.: Condens. Matter* **2009**, *21*, 474221.

(30) Sambirski, E. J.; Schwartz, D. C.; de Pablo, J. J. *Biophys. J.* **2009**, *96*, 1675−1690.

(31) Sambirski, E. J.; Ortiz, V.; de Pablo, J. J. *J. Phys.: Condens. Matter* **2009**, *21*, 034105.

(32) Ouldridge, T. E.; Johnston, I. G.; Louis, A. A.; Doye, J. P. *J. Chem. Phys.* **2009**, *130*, 065101.

(33) Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. *J. Chem. Theory Comput.* **2010**, *6*, 1711−1725.

(34) Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S. *J. Chem. Phys.* **2010**, *132*, 035105.

(35) Linak, M. C.; Dorfman, K. D. *J. Chem. Phys.* **2010**, *133*, 125101.

(36) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. *Phys. Rev. Lett.* **2010**, *104*, 178101.

(37) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. *J. Chem. Phys.* **2011**, *134*, 085101.

(38) Freeman, G. S.; Hinckley, D. M.; de Pablo, J. J. *J. Chem. Phys.* **2011**, *135*, 165104.

(39) Linak, M. C.; Tourdot, R.; Dorfman, K. D. *J. Chem. Phys.* **2011**, *135*, 205102.

(40) Hsu, C. W.; Fyta, M.; Lakatos, G.; Melchionna, S.; Kaxiras, E. *J. Chem. Phys.* **2012**, *137*, 105102.

(41) Edens, L. E.; Brozik, J. A.; Keller, D. J. *J. Phys. Chem. B* **2012**, *116*, 14735−14743.

(42) Gonzalez, O.; Petkevičiūtė, D.; Maddocks, J. H. *J. Chem. Phys.* **2013**, *138*, 055102.

(43) He, Y.; Maciejczyk, M.; Oldziej, S.; Scheraga, H. A.; Liwo, A. *Phys. Rev. Lett.* **2013**, *110*, 098101.

(44) Ouldridge, T. E.; Šulc, P.; Romano, F.; Doye, P. K. J.; Louis, A. A. *Nucleic Acids Res.* **2013**, *41*, 8886−8895.

(45) Cragnolini, T.; Derreumaux, P.; Pasquali, S. *J. Phys. Chem. B* **2013**, *117*, 8047−8060.

(46) Hoang, T. X.; Cieplak, M. *J. Chem. Phys.* **2000**, *112*, 6851−6862.

(47) Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. A. *J. Comput. Chem.* **2010**, *31*, 1644−1655.

(48) Leimkuhler, B. J.; Reich, S. *Simulating Hamiltonian Dynamics*; Cambridge University Press: Cambridge, U.K., 2004; pp 199−233.

(49) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24−34.

(50) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(51) Kol, A.; Laird, B. B.; Leimkuhler, B. J. *J. Chem. Phys.* **1997**, *107*, 2580−2588.

(52) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323−2347.

(53) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1715−1731.

(54) Tsonchev, S.; Schatz, G. C.; Ratner, M. A. *Chem. Phys. Lett.* **2004**, *400*, 221−225.

(55) Mazur, J.; Jernigan, R. L. *Biopolymers* **1991**, *31*, 1615−1629.

(56) Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide*; Springer: New York, 2010; p 177.

(57) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(58) Cheatham, T. E., III; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.

(59) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817−3829.

(60) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(61) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339−1350.

(62) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275−278.

(63) Grossfield, A. *WHAM: the weighted histogram analysis method*, http://membrane.urmc.rochester.edu/content/wham (accessed June 1, 2012).

(64) Boggs, P. T.; Byrd, R. H.; Schnabel, R. B. *Numerical Optimization 1984: Proceedings of the Siam Conference on Numerical Optimization*, Boulder, Colorado; 1984; pp 71−83.

(65) Arnott, S.; Hukins, D. W. L. *Biochem. Bioph. Res. Commun.* **1972**, *47*, 1504−1509.

(66) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(67) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671−680.

(68) Černý, V. *J. Opt. Theory Appl.* **1985**, *45*, 41−51.

(69) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes. The Art of Scientific Computing*; Cambridge University Press: New York, 2007; pp 509−514.

(70) Doi, M.; Edwards, S. F. *The Theory of Polymer Dynamics*; Oxford University Press: New York, 1988; pp 316−317.

(71) Fathizadeh, A.; Eslami-Mossollam, B.; Ejtehadi, M. R. *Phys. Rev. E* **2012**, *86*, 051907.

(72) Quinn, M. J. *Parallel Programming in C with MPI and OpenMP*; McGraw-Hill Companies: New York, 2004; pp 404−435.

(73) Tinland, B.; Pluen, A.; Sturm, J.; Weill, G. *Macromolecules* **1997**, *30*, 5763−5765.

(74) Bustamante, C.; Marko, J. F.; Siggia, E. D.; Smith, S. *Science* **1994**, *265*, 1599−1600.

(75) Smith, S. B.; Cui, Y.; Bustamante, C. *Science* **1996**, *271*, 795−799.

(76) Baumann, C. G.; Smith, S. B.; Boomfield, V. A.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6185−6190.

(77) Geggier, S.; Kotlyar, A.; Vologodskii, A. *Nucleic Acids Res.* **2011**, *4*, 1419−1426.