An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization

Daniel M. Hinckley,[1] Gordon S. Freeman,[1] Jonathan K. Whitmer,[2] and Juan J. de Pablo[2,3,a)]

[1] *Department of Chemical and Biological Engineering, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA*
[2] *Institute for Molecular Engineering, Argonne National Laboratory, Chicago, Illinois 60439, USA*
[3] *Institute for Molecular Engineering, University of Chicago, Chicago, Illinois 60637, USA*

(Received 24 January 2013; accepted 25 August 2013; published online 7 October 2013)

A new **3**-**S**ite-**P**er-**N**ucleotide coarse-grained model for DNA is presented. The model includes anisotropic potentials between bases involved in base stacking and base pair interactions that enable the description of relevant structural properties, including the major and minor grooves. In an improvement over available coarse-grained models, the correct persistence length is recovered for both ssDNA and dsDNA, allowing for simulation of non-canonical structures such as hairpins. DNA melting temperatures, measured for duplexes and hairpins by integrating over free energy surfaces generated using metadynamics simulations, are shown to be in quantitative agreement with experiment for a variety of sequences and conditions. Hybridization rate constants, calculated using forward-flux sampling, are also shown to be in good agreement with experiment. The coarse-grained model presented here is suitable for use in biological and engineering applications, including nucleosome positioning and DNA-templated engineering. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4822042]

I. INTRODUCTION

The elementary building block of DNA consists of a sugar, a phosphate, and a base. The bases can be Adenine (A), Thymine (T), Guanine (G) or Cytosine (C), which differ primarily in the number of aromatic rings and their functional groups. The small number of nucleic acid monomers and the similarity between them stand in sharp contrast to the diverse assortment of amino acids that constitute proteins. Such relative homogeneity makes DNA particularly amenable to coarse-grained descriptions.

There is considerable interest in developing molecular models capable of describing the structure and properties of DNA. In order to complement available experimental information, such models must provide appropriate temporal and spatial resolution. Existing atomistic force fields are able to describe fast conformational fluctuations of DNA structure and protein-DNA binding;[1] however, such models cannot access the length and time scales needed to study phenomena such as nucleosome positioning or DNA hybridization. Coarse-grained models that reduce the total number of degrees of freedom are therefore necessary.

Examples of DNA coarse-grained models include representations that rely on rigid rods, semi-flexible rods,[2,3] or bead-spring chains[4,5] with different spring potentials.[6] Higher resolution n-site-per-nucleotide models have been proposed in attempts to resolve more of the molecular structure of DNA. Beginning with an $n = 2$ representation, coarse-grained models of increasing complexity have been introduced.[7–22] By increasing n, it has become possible to reproduce more of the structural features of DNA, such as the major and minor grooves, as well as thermodynamic properties such as the melting temperature.[9,14,17,18,22]

Here we focus on the so-called **3**-**S**ite-**P**er-**N**ucleotide (3SPN) model, originally introduced by Knotts *et al.*[9] Note that $n = 3$ is the minimum number of sites that enable resolution of major and minor grooves using isotropic potentials, thereby providing a reasonable compromise between structural fidelity and computational efficiency. 3SPN.1, a subsequent refinement of the model, enabled a description of denaturation and renaturation as a function of temperature and ionic strength.[11] This model has been used to explore the pathways for hybridization in the bulk[23,24] and on surfaces,[25–27] and has provided new insights into DNA bending,[28] DNA melting,[29] and DNA-based hybrid structures.[30] It has since been extended to include explicit ions[30–32] and water.[31]

However, available versions of the 3SPN model exhibit a number of important limitations. In particular, the persistence length of dsDNA was underpredicted[9] or the persistence length of ssDNA was overpredicted.[11] As pointed out by us[9,11] and others,[16] the use of Gō-like interactions to represent base stacking restricts the range of conformations that may be sampled by the coarse-grained model. Gō-like interactions use simple non-bonded potentials to penalize deviations from a reference structure.[33] These interactions were adopted in previous models because only one site, subject to isotropic potentials, was used to represent each base; without additional constraints the duplex would be neither stable nor helical.

Other limitations include the use of isotropic base pairing interactions, which allow a single base to pair with multiple bases on the complementary strand.[28,29] The simulated melting temperatures did not include concentration effects[34]

[a)] Electronic mail: depablo@uchicago.edu and were dependent on a rather arbitrary definition of a base pair.[14] Perhaps more importantly, the hybridization rate constants calculated with 3SPN.1 deviate significantly from experimental values, and interactions between 3SPN.1 and proteins can lead to structural motifs that are not consistent with experimental observations.

Previous attempts to improve 3SPN include the work of Morriss-Andrews *et al.*,[16] who developed a three-site model with base sites represented by ellipsoidal particles. The asymmetrical nature of the ellipsoids removed the need for Gō-like interactions, but unfortunately led to structural and mechanical characteristics that are not in agreement with experiment.

Following a different approach, Dans *et al.*[14] chose to incorporate additional interaction sites into the model; they adopted a 6-site-per-nucleotide representation with explicit partial charges, thereby providing a simple means to map all-atom (AA) coordinates onto the CG model.[35] This model has been shown to provide reasonable agreement with base step parameters from AA simulations; however, it includes multimodal torsional interactions that could impart unrealistic rigidity to ssDNA.

Ouldridge *et al.*[17] developed a model with 3 collinear sites and a vector normal to the base site. This normal vector allowed for the construction of angle-dependent potential interactions, which can in turn reproduce the persistence length of both ssDNA and dsDNA. The model can also describe the ssDNA–dsDNA transition, and a variety of DNA structures.[36–38] The original model had limited sequence-specificity and was developed in the context of a Monte Carlo framework, but a more recent version includes sequence-specificity, and a Langevin dynamics framework has been built around it.[39] This model, however, does not capture the major and minor grooves of DNA, thereby limiting its applicability to the study of DNA-protein interactions. Linak *et al.*[18] also incorporated angle-dependent base pairing, base, and cross-stacking interactions into a 3SPN model to simulate hairpins, G-quartets, and triplexes. However, some anomalies in the model have become apparent and must be addressed before it can be used more broadly.[40]

Hsu *et al.*[20] developed a two site coarse-grained DNA model that is based on *ab initio* calculations. Using density functional theory, they arrived at a set of parameters for bonded and angle-dependent non-bonded interactions. The resulting model is able to resolve the correct DNA structure and qualitatively captures bubble formation within long helices. However, it overpredicts the melting temperature of DNA considerably, making it of limited use for study of structural transitions. That model does not provide a good representation of the excluded volume of DNA, and cannot be used in its present form to study DNA-protein complexes.

Recently, several approaches have been proposed that provide more detailed representations of the nucleotide bases. Edens *et al.*[21] developed a coarse-grained model with a rigid nucleotide consisting of 7 or 8 sites designed to create a lock-and-key relationship with the complementary nucleotide. By creating a close-packed representation of the excluded volume core, it reproduces the major and minor grooves while utilizing short-ranged isotropic potentials. Despite not being rigorously parameterized to capture thermodynamic quantities, it captures the structure and persistence length of DNA and has been used to examine the effect of over- and undertwist on DNA minicircles. Taking the representation of the base to an even finer level, Savin *et al.*[19] developed a 6-site-per-nucleotide representation of DNA that facilitates inter-conversion into AA coordinates. This allows for base stacking and base pairing interactions to be calculated at every time step using the AMBER AA force field.[41] This model has been used to calculate the thermal conductivity of DNA[19] and to explain the co-existence of multiple phases within stretched dsDNA.[42] However, it is unknown to what degree it can describe the thermodynamic properties of DNA. More recently, He *et al.*[22] proposed a coarse-grained model that couples a structured backbone with bases represented by Gay-Berne ellipsoids and electrostatic dipoles. The model was parameterized using AA simulations and distributions extracted from PDB structures, and its results suggest that multipole-multipole interactions are important in driving the formation of a DNA double helix.

All three models involve significant departures from previous descriptions of DNA, and rely on the inclusion of additional interaction sites or higher order interactions. In the present work, we seek to develop a model with a minimal number of interaction sites and with interaction potentials that are easily implemented into existing molecular dynamics packages. As in previous versions of 3SPN, we place emphasis on the description of the equilibrium structure and thermodynamic properties of DNA.

Generally speaking, the success of available coarse-grained DNA models serves to highlight the usefulness of a top-down parameterization approach[9,11,17,28] that relies on available experimentally-measured quantities, thereby leading to greater applicability. Despite important bottom-up efforts to parameterize coarse-grained models by relying almost exclusively on AA simulations,[16] or using data extracted from nucleic acids in structural databases,[22] the predictive capability of the resulting models has been limited. Here we note, however, that recent work suggests that, by including higher order terms, bottom-up parameterizations of DNA can become more effective.[19,20,22]

In this work, we follow a similar top-down strategy to that adopted in our earlier work[9,11] to develop an improved 3SPN model that eliminates many of the shortcomings of previous versions. The experimental data employed in our work include: base step energies,[43] base stacking free energies,[44] and equilibrium values of bond lengths, bend angles, and dihedral angles.[45] We demonstrate that by including such information, it is possible to build on available models and extend their predictive capability to a wider range of experimental situations. More specifically, the improved 3SPN coarse-grained model for DNA (1) resolves the correct structural features, (2) accurately reproduces the persistence lengths of both ss- and ds-DNA, (3) incorporates the effect of ionic strength, sequence, concentration, and temperature on hairpin and duplex formation, and (4) predicts reaction rate constants that are consistent with experimental values.

This paper is structured as follows: In Sec. II, we present the parameterization, topology, and Hamiltonian of our new model, which we refer to as "3SPN.2." In Sec. III we describe methods used to calculate structural properties, persistence lengths, melting temperatures, and hybridization rate constants. In Sec. IV we compare the predictions from 3SPN.2 to experimental data and to those of an earlier version of the model (3SPN.1).[11] Lastly, In Sec. V we discuss the improvements, applications and limitations of the 3SPN.2 model.

II. MODEL

A. Parameterization approach

We seek to parameterize coarse-grained interactions in a manner that ensures consistency with experimental free energy measures. Specifically, we target free energies of hybridization and base stacking.

The free energy of hybridization is directly related to the melting temperature T_m of a DNA sequence. At T_m, the relative free energies (and probabilities) of the hybridized and dehybridized states are equal. By adjusting model parameters such that the free energies of each state are equal at T_m, we can determine the proper strength for inter-strand non-bonded interactions. Values of T_m, obtained via UV absorbance measurements, are widely available; here we use data from Owczarzy *et al.*[46]

The intra-strand base stacking was measured by Protozanova *et al.* by nicking DNA duplexes and then examining their relative electrophoretic mobility.[44] These free energies can be interpreted as the relative probabilities of the stacked and destacked states. In simulations, we can calculate these probabilities and adjust stacking strengths until the simulated free energy of stacking is consistent with experiment. We do so by relying on a Boltzmann inversion approach.[47] By doing so, we assume that the additional restraint of backbone connectivity has a negligible effect on the required strength of base stacking interactions.

Hybridization and stacking free energies are calculated by metadynamics simulations.[48] In these simulations, a time-dependent bias is added to the system, forcing the simulation to sample all relevant regions of an order parameter space. Upon convergence, all states are sampled uniformly and the bias provides an estimate of the free energy as a function of the order parameter(s). By integrating the free energy surface over relevant ranges of the order parameters, the relatively probability of each state can be calculated. Additional details regarding these calculations can be found in Sec. III C and in the supplementary material.[49]

Structural quantities, such as the persistence length or the width of the minor and major grooves, are obtained by adjusting bonded parameters manually until results are consistent with experiment. Due to the indirect effect of the bonded parameters on the aforementioned free energies, several iterations are performed until consistency is reached with both thermodynamic and structural properties.

B. Site diameter

Nucleotides are represented by 3 spherical sites corresponding to the phosphate, deoxyribose sugar, and nitrogenous base—these are placed at the center of mass (COM) of the corresponding moiety. This differs from prior versions of 3SPN[9,11] that placed the base site at the N1 atom site of Adenine and Guanine and the N3 atom site of Thymine and Cytosine. This change is motivated by the close stacking of bases in DNA that excludes ions and water molecules from the core of the double helix. 3SPN.1 represented each base site with the same small excluded volume ($\sigma = 6.86$ Å) at locations on the very edge of each base (the N1 or N3 atoms). This resulted in an excluded volume representation that would permit ions or coarse-grained protein sites to insert themselves into the DNA. By centering the base at its COM, we more accurately represent the solvent excluded core of the DNA. This will be beneficial in future implementations of 3SPN.2 with explicit ions.

The size of each base site is set such that no excluded volume interactions occur in coarse-grained representations of the B form of DNA (B-DNA). The resulting site diameters run contrary to intuition (e.g., the A and G sites are smaller than the T and C sites); this is a consequence of the use of an isotropic excluded volume potential to represent the heterocyclic bases, which are anisotropic in nature. However, this choice does not limit the applicability of the model to studies of DNA-protein binding. The resulting model is depicted in Fig. 1. The numerical values of these diameters and a more extensive discussion of their origin are given in Appendix A.

C. Bonded potentials

The bonded potential, U_b, includes bond, angle, and dihedral contributions. The bond contributions are harmonic and anharmonic. The angle contribution is harmonic, as in

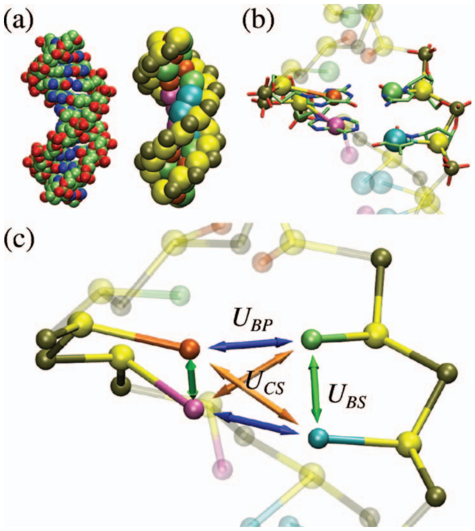

FIG. 1. (a) Comparison of all-atom and 3SPN.2 coarse-grained excluded volume representations of the Drew-Dickerson dodecamer (5′-CGCGAATTCGCG-3′). (b) Coarse-grained sites of 3SPN.2 superimposed on the all-atom representations. The coarse-grained sites are located at the centers of mass of the phosphate, sugar, or base. (c) Schematic representation of the angle-dependent non-bonded interactions acting between the base sites. The green arrows represent the base stacking potential U_{BS}, the blue arrows represent the base pairing potential U_{BP}, and the orange arrows represent the cross-stacking potential U_{CS}. Figures were rendered using VMD.[50]

previous versions[9,11] of 3SPN. The dihedral potential is given by a Gaussian well. Thus, the bonded potential is

$$\begin{aligned} U_{\mathrm{b}} &= U_{\mathrm{bond}} + U_{\mathrm{bend}} + U_{\mathrm{tors}} \\ &= \sum_i^{\mathrm{bonds}} k_b(r_i - r_{o,i})^2 + 100k_b(r_i - r_{o,i})^4 \\ &\quad + \sum_i^{\mathrm{bends}} k_\theta(\theta_i - \theta_{o,i})^2 \\ &\quad + \sum_i^{\mathrm{dihedrals}} -k_\phi \exp\left(\frac{-(\phi_i - \phi_{o,i})^2}{2\sigma_{\phi,i}^2}\right), \end{aligned} \tag{1}$$

where k_b and $r_{o,\,i}$ are the force constant and equilibrium bond length for bond i; k_θ and $\theta_{o,\,i}$ represent the force constant and equilibrium angle for bend i, and k_ϕ, $\phi_{o,\,i}$, and $\sigma_{\phi,\,i}$ denote the well-depth, equilibrium angle, and Gaussian well-width, respectively, of dihedral i. The functional form of the dihedral potential represents a departure from previous versions of the 3SPN model; it was changed to a Gaussian well in order to favor B-DNA while still allowing free rotation once the structure is perturbed. While not strictly periodic, the potential and its first derivative effectively go to zero at $|\phi_i - \phi_{o,\,i}| = \pi$, allowing a periodic implementation. In addition, 3SPN.2 only applies a dihedral potential between dihedrals formed by sugar and phosphates sites (i.e., S-P-S-P and P-S-P-S). Previous versions of 3SPN included dihedral potentials between dihedrals involving bases, such as A-S-P-S, S-P-S-A, etc. These are no longer necessary because of the anisotropic nature of the non-bonded interactions, which is discussed in Sec. II D. All equilibrium bond lengths $r_{o,\,i}$, bend angles $\theta_{o,\,i}$, and dihedral angles $\phi_{o,\,i}$ are obtained from the fiber crystal structure of B-DNA.[45]

D. Non-bonded potentials

The non-bonded potential U_{nb} is given by

$$U_{\mathrm{nb}} = U_{\mathrm{exe}} + U_{\mathrm{bstk}} + U_{\mathrm{cstk}} + U_{\mathrm{bp}} + U_{\mathrm{elec}}, \tag{2}$$

where U_{exe} denotes excluded volume contributions and U_{bstk}, U_{cstk}, and U_{bp} are the intra-strand base stacking, inter-strand cross-stacking, and base pairing interactions, respectively. The term U_{elec} is a screened electrostatic potential. It is important to note that these non-bonded contributions to the energy arise only between sites that do not participate in the same bond, angle, or dihedral potential.

1. Excluded volume interactions

We utilize a purely repulsive potential between sites i and j of the form

$$U_{\mathrm{exe}} = \sum_{i<j} \begin{cases} \epsilon_{\mathrm{r}}\left[\left(\dfrac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\dfrac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \epsilon_{\mathrm{r}} & r < r_C \\ 0 & r \geq r_C \end{cases}. \tag{3}$$

This potential is active between all sites that do not participate in bonded interactions, base pairing interactions (U_{bp}), or are not part of neighboring nucleotides in the same strand and are located within a cutoff distance r_C. The energy parameter for excluded volume interactions is denoted by ϵ_{r}; $\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)$ is the average site diameter, and r_{ij} is the intersite separation. The cutoff distance r_C is always σ_{ij}. Bases that form W–C base pairs do not interact via this potential; instead, they experience a repulsion dictated by a Morse potential as explained in Sec. II D 2.

2. Base–base interactions

As mentioned earlier, prior versions of 3SPN relied on an isotropic intra-strand Gō-like interaction[33] to capture base stacking and stabilize dsDNA. These interactions, originally developed in the context of studies of protein folding, allow for fluctuations around a reference native structure while permitting unfolding in denaturing conditions. A disadvantage is that the resulting model is unable to capture metastable states that deviate strongly from the reference structure.

Prior versions of 3SPN also used an isotropic potential to represent base pairing interactions, and were unable to account for the directionality intrinsic in W–C base pairing. In 3SPN.2 the Gō-like interactions are replaced by angle-dependent potentials, as done in other models.[17,18,20,22] This angle-dependence creates a cone of strong attraction surrounded by a larger cone within which a Morse potential is smoothly modulated from its full magnitude to zero. Outside of the larger cone, there is no interaction other than the excluded volume repulsion. This modulation is illustrated in Fig. 2.

The angles that modulate the potentials are shown in Fig. 3. The angle used to modulate intra-strand base stacking, θ_{BS}, arises between a vector connecting a sugar and base site within the same nucleotide and the vector connecting the current base with its neighbor in the 5′ or 3′ direction.

For base pairing, the angles θ_1 and θ_2 are used to modulate the Morse potential. The first of these, θ_1, is the angle between the vector pointing from the sugar site to the base site in a nucleotide participating in a W–C base pair and the vector connecting the two complementary bases; θ_2 is the angle between the vector connecting the two base sites participating

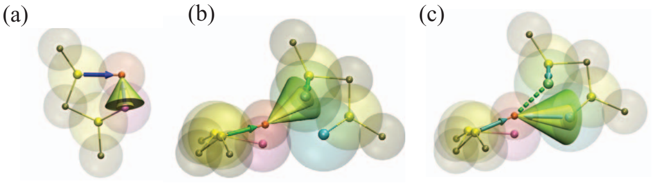

FIG. 2. Schematic representation of anisotropic base-base interactions. The inner yellow cone represents the angles wherein the full potential is applied. The volume of the outer green cone not included in the yellow cone represents the range of angles wherein the potential is modulated from its full value to zero. A parameter K controls the width of these cones, with a smaller value leading to a wider range of interactions. (a) Angle-dependence of the intra-strand base stacking interactions ($K = 6$). (b) Angle-dependence of one of the angles θ in the base pairing interactions ($K = 12$). (c) Angle-dependence of the cross-stacking interactions ($K = 8$). The shaded spheres represent the excluded volume of the DNA sites. Figures were rendered using VMD.[50]

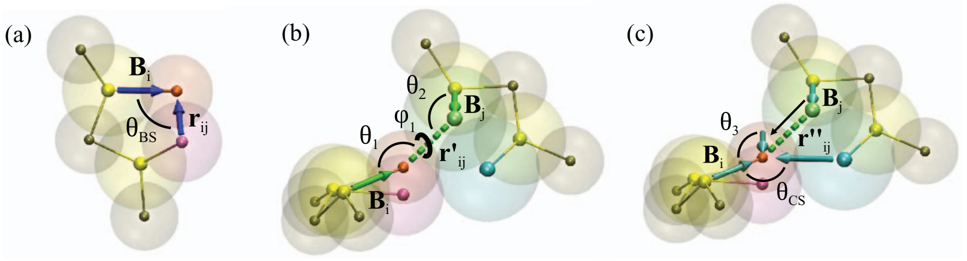

FIG. 3. Definition of the angles used to modulate base-base interactions. (a) Intra-strand base stacking angle θ_{BS}, defined as the angle between $\mathbf{B}_i$, the vector connecting a sugar and base site and $\mathbf{r}_{ij}$, the vector connecting the base with its neighbor in the 3′ direction. (b) Base pairing angles, θ_1, θ_2, and ϕ_1, are defined by the effective dihedral angle between the base pair; θ_1 is the angle between vector $\mathbf{B}_i$ connecting the sugar and base on the sense strand and vector $\mathbf{r}'_{ij}$ connecting the base on the sense strand with its complement, either on the anti-sense strand in the case of a duplex or the sense strand of a hairpin. Here θ_2 is the angle between vector $\mathbf{r}'_{ij}$ and vector $\mathbf{B}_j$, connecting the sugar and base site of the complementary nucleotide. The angle ϕ_1 is the dihedral angle defined by the vectors $\mathbf{B}_i$, $\mathbf{r}_{ij}$, and $\mathbf{B}_j$. (c) cross-stacking angles, θ_3 and θ_{CS}, are the angles between vectors $\mathbf{B}_i$ and $\mathbf{B}_j$ and vectors $\mathbf{B}_i$ and $\mathbf{r}''_{ij}$, respectively. Here, $\mathbf{r}''_{ij}$ is the vector connecting the base site on the sense strand to the base site adjacent to its W–C complement in the 5′ direction of the anti-sense strand. A similar vector is defined between the base site participating in a W–C base pair and the base site adjacent to its complement in the 3′ direction of the sense strand. The shaded spheres represent the excluded volume of the DNA sites. Figures were rendered using VMD.[50]

in a base pair and the sugar–base vector on the complementary nucleotide, and ϕ_1 is the dihedral defined by the 3 vectors used to define θ_1 and θ_2. Note that the dihedral angle is not used in the modulating function.

Two angles modulate the cross-stacking interactions: θ_3 is the angle between the sugar–base vectors on sites participating in a W–C base pair, and θ_{CS} is the angle between one of the sugar–base vectors and a vector connecting that same base site to the base site immediately adjacent (3′ direction) to the W–C complement on the opposite strand. As such, these angles are similar to those introduced in the work of Linak *et al.*;[18] our treatment differs in that the equilibrium angles for each interaction are unique to the identity of the bases involved, and in each case have been extracted from the crystal structure of B-DNA.

The modulating function *f* takes the form

$$f(K,\Delta\theta) = \begin{cases} 1 & \frac{-\pi}{2K} < \Delta\theta < \frac{\pi}{2K} \\ 1-\cos^2(K\Delta\theta) & -\frac{\pi}{K} < \Delta\theta < -\frac{\pi}{2K} \text{ or } \frac{\pi}{2K} < \Delta\theta < \frac{\pi}{K} \\ 0 & \Delta\theta < -\frac{\pi}{K} \text{ or } \Delta\theta > \frac{\pi}{K} \end{cases}, \tag{4}$$

where K controls the width of the cone of attraction and $\Delta\theta = \theta - \theta_o$, the difference between the current angle θ and θ_o, the angle from the crystal structure of B-DNA. The derivative of this function vanishes at $|\frac{\pi}{K}|$ and $|\frac{\pi}{2K}|$. The relative widths of these modulating regions are shown in Fig. 2.

Base stacking and base pairing interactions between base sites are described by a Morse potential,

$$U_{\mathrm{m}}(\epsilon_{ij}, \alpha_{ij}, r_{ij}) = \epsilon_{ij}(1 - e^{(-\alpha_{ij}(r_{ij}-r_{o,ij}))})^2 - \epsilon_{ij}, \tag{5}$$

where ϵ_{ij} is the depth of the well of the attraction between sites i and j, $r_{o,\,ij}$ is the equilibrium separation between them, and α_{ij} is a parameter that is adjusted to control the range of attraction. In the course of model development, we have found it helpful to decompose the attractive and repulsive portions of the Morse potential when modulating the attraction using *f*. After decomposition, the repulsive component is

$$U_{\mathrm{m}}^{\mathrm{rep}}(\epsilon_{ij}, \alpha_{ij}, r_{ij}) = \begin{cases} \epsilon_{ij}(1 - e^{(-\alpha_{ij}(r_{ij}-r_{o,ij}))})^2 & r < r_{o,ij} \\ 0 & r \geq r_{o,ij} \end{cases} \tag{6}$$

and the attractive component is

$$U_{\mathrm{m}}^{\mathrm{attr}}(\epsilon_{ij}, \alpha_{ij}, r_{ij}) = \begin{cases} -\epsilon_{ij} & r_{ij} < r_{o,ij} \\ \epsilon_{ij}(1 - e^{(-\alpha_{ij}(r_{ij}-r_{o,ij}))})^2 - \epsilon_{ij} & r_{ij} \geq r_{o,ij}. \end{cases} \tag{7}$$

This decomposition ensures that the repulsive character is maintained, regardless of the modulating angle.

Base pairing and base stacking interactions are modulated by *f* using θ_{BS} and θ_1 and θ_2, respectively. The cross-stacking interaction is modulated using both θ_3 and θ_{CS}. Only the attractive component of the Morse potential is modulated. The resulting functional form is given by

$$U_{\mathrm{bstk}} = \sum^{n_{\mathrm{bstk}}} \begin{cases} U_{\mathrm{m}}^{\mathrm{rep}}(\epsilon_{ij}, \alpha_{\mathrm{BS}}, r_{ij}) + f(K_{\mathrm{BS}}, \Delta\theta_{\mathrm{BS}ij}) U_{\mathrm{m}}^{\mathrm{attr}}(\epsilon_{ij}, \alpha_{\mathrm{BS}}, r_{ij}) & r_{ij} < r_{o,ij} \\ f(K_{\mathrm{BS}}, \Delta\theta_{\mathrm{BS}ij}) U_{\mathrm{m}}^{\mathrm{attr}}(\epsilon_{ij}, \alpha_{\mathrm{BS}}, r_{ij}) & r_{ij} \geq r_{o,ij} \end{cases} \tag{8}$$

for base stacking,

$$U_{\mathrm{bp}} = \sum^{n_{\mathrm{bp}}} \begin{cases} U_{\mathrm{m}}^{\mathrm{rep}}(\epsilon_{ij}, \alpha_{\mathrm{BP}}, r_{ij}) + \frac{1}{2}(1+\cos(\Delta\phi_1)) f(K_{\mathrm{BP}}, \Delta\theta_{1_{ij}}) f(K_{\mathrm{BP}}, \Delta\theta_{2ij}) U_{\mathrm{m}}^{\mathrm{attr}}(\epsilon_{ij}, \alpha_{\mathrm{BP}}, r_{ij}) & r_{ij} < r_{o,ij} \\ \frac{1}{2}(1+\cos(\Delta\phi_1)) f(K_{\mathrm{BP}}, \Delta\theta_{1ij}) f(K_{\mathrm{BP}}, \Delta\theta_{2ij}) U_{\mathrm{m}}^{\mathrm{attr}}(\epsilon_{ij}, \alpha_{\mathrm{BP}}, r_{ij}) & r_{ij} \geq r_{o,ij} \end{cases} \tag{9}$$

for base pairing, and

$$U_{\text{cstk}} = \sum^{n_{\text{cstk}}} f(K_{\text{BP}}, \Delta\theta_{3ij}))f(K_{\text{CS}}, \Delta\theta_{\text{CS}ij})U_{\text{m}}^{\text{attr}}(\epsilon_{ij}, \alpha_{\text{CS}}, r_{ij}) \tag{10}$$

for cross-stacking. In the equations above, ϵ_{ij} and $r_{o,\,ij}$ are specific to each combination of bases i and j; $\Delta\phi_1 = \phi_1 - \phi_{1_o}$, penalizing deviations from a reference dihedral angle. The integers n_{bstk}, n_{bp} and n_{cstk} represent the total number of base stacking, base pairing, and cross-stacking interactions, respectively. The parameters K_{BS} and α_{BS}, K_{BP} and α_{BP}, K_{CS} and α_{CS} are the same for all base stacking, base pairing, and cross-stacking interactions, and are given in Appendix B. We restrict intra-strand base pairing to occur only between base sites separated by at least 3 nucleotides.

The functional form of the base pairing potential is designed to prevent the formation of multiple base pairs, as observed in previous models.[28,29] By using the effective dihedral created by the W–C pair and the neighboring sugar sites, we discourage the formation of multiple base pairs by a single base, as might otherwise be seen in an AA base step.

E. Electrostatic interactions

Electrostatic interactions occur between the charged phosphates sites; the base and sugar sites are neutral. In a departure from previous versions of the 3SPN model, the effective charge of each phosphate is increased from -1.0 to -0.6, giving each nucleotide a net charge of -0.6. This increase in the effective charge is motivated by Oosawa–Manning counter-ion condensation theory,[51,52] which states that the charge density along a polyelectrolyte will be reduced through counter-ion condensation until the electrostatic interactions within the chain are comparable to k_bT, the thermal energy. For a polyelectrolyte with inter-charge spacing b, counter-ion condensation will occur if $\frac{\lambda_B}{b} > 1$. Here λ_B is the Bjerrum length:

$$\lambda_B = \frac{e_c^2}{4\pi\epsilon_o\epsilon k_b T}, \tag{11}$$

where e_c is the elementary charge, ϵ_o is the dielectric permittivity of vacuum, ϵ is the dielectric permittivity of the solution (defined below), k_b is the Boltzmann constant, and T is the temperature. The Bjerrum length λ_B for water at 298 K is 7.14 Å while the approximate charge spacing in ssDNA is 4.3 Å;[53,54] therefore, counter-ion condensation will occur. We obtain the reduced charge of -0.6 by rounding $\frac{-b}{\lambda_B}$. We use the value b for ssDNA instead of dsDNA because electrostatic interactions have a dominant effect on the self-assembly of ssDNA molecules into dsDNA. The approximate charge spacing for dsDNA is $-2e^-/3.4$ Å and would lead to an effective charge -0.2 to -0.25.

While Manning's assumption that the polyelectrolyte consists of an infinite linear charge density is not consistent with the flexible nature of ssDNA, Liu and Muthukumar have shown[55] that for monovalent cations, Oosawa-Manning theory provides a good approximation for flexible polymers, thereby suggesting that the theory can be applied to our coarse-grained model.

The charged phosphate sites interact via a screened electrostatic potential between all other inter-strand phosphates and all intra-strand phosphates not on neighboring nucleotides. This interaction is given by

$$U_{\text{elec}} = \sum_{i<j}^{n_{elec}} \frac{q_i q_j e^{-r_{ij}/\lambda_D}}{4\pi\epsilon_o\epsilon(T,C)r_{ij}}, \tag{12}$$

where q_i and q_j are the charges of sites i and j, r_{ij} is the intersite separation, λ_D is the Debye screening length, $\epsilon(T, C)$ is the dielectric permittivity of the solution, and other variables as defined previously. The Debye screening length is defined as

$$\lambda_D = \sqrt{\frac{\epsilon_o\epsilon(T,C)}{2\beta N_A e_c^2 I}}, \tag{13}$$

where β is the inverse thermal energy of the system $(k_bT)^{-1}$, N_A is Avogadro's number, and I is the ionic strength of the solution. The solution dielectric permittivity $\epsilon(T, C)$ is a function of the molarity of NaCl and temperature. Assuming that the contributions of salt molarity C and temperature T are independent, as done by Ref. 56, the dielectric permittivity can be approximated as

$$\epsilon(T,C) = \epsilon(T)a(C), \tag{14}$$

where

$$\epsilon(T) = 249.4 - 0.788T/K + 7.20 \times 10^{-4}(T/K)^2 \tag{15}$$

and

$$\begin{aligned} a(C) =& 1.000 - 2.551C/M + 5.151 \\ & \times 10^{-2}(C/M)^2 - 6.889 \times 10^{-3}(C/M)^3. \end{aligned} \tag{16}$$

Equation (15) is as reported in Ref. 56, while Eq. (16) has a form inspired by others[57] and coefficients obtained by fitting experimental data.[58]

F. Langevin dynamics

The integration scheme used here is the same as in previous versions of 3SPN[9,11] and relies on Langevin dynamics (LD). The force on particle i at point r is given by

$$f_i(r) = -\nabla U_i(r) - \gamma_i p_i(t) + g_i(t) = \frac{dp_i(t)}{dt}, \tag{17}$$

where $-\nabla U_i(r)$ is the force arising from the force field, γ_i is a damping constant (in units of reciprocal time), $p_i(t)$ is the momentum of particle i, and $g_i(t)$ is a random force that satisfies the fluctuation-dissipation theorem. This LD algorithm is implemented as reported by Bussi and Parinello[59] with a time step Δt of 0.02 ps.

In order to assign the damping constant, we use the Einstein relation which states

$$D = \frac{1}{\beta\xi_m}, \tag{18}$$

where ξ_m is the molecular friction coefficient. We assume that the molecular friction coefficient is distributed uniformly among all N coarse-grained DNA sites as in 3SPN.1,[11] irrespective of the diameter of the different sites. This gives

$$D = \frac{1}{\beta N \xi_i}, \tag{19}$$

where ξ_i is the friction coefficient of each site. We use an experimental diffusivity of an 18 base pair (bp) ssDNA ($9.94\ 10^{-7}$ cm^2/s) and scale the diffusivity according to $D \sim 1/N^{0.68}$ to account for the effect of length.[60] Having calculated ξ_i, the damping constant is simply $\gamma_i = \frac{\xi_i}{m_i}$.

III. VALIDATION METHODS

A. Structural properties

Relevant structural properties of DNA include the width of the duplex, the number of base pairs per turn, the base rise, and the widths of the major and minor grooves. We calculate the width of dsDNA from long-time averages of the distances between phosphate sites of a W–C base pair. These bases are located at least 5 bases from the end of the strand to avoid end effects (e.g., fraying). Base rise is calculated by projecting the inter-base separation onto an approximate helical axis, as constructed below. Sequence and groove geometry determine the relative affinity of a protein to a particular sequence of DNA.[54] Previous versions of 3SPN did not exhibit stable, clearly defined major and minor grooves, making such models ill-suited for studying DNA–protein binding events.

In order to calculate the width of the major and minor grooves, we use the method outlined by Stofer and Lavery,[61] where a plane is rotated around a vector perpendicular to the helical axis that bisects the minor groove. The distance between the points of intersection of this plane with continuous space curves drawn through the phosphates of each strand is calculated for different angles of rotation. The local minima in the resulting curve showing distance as a function of rotation angle represent the major and minor groove widths.

When using this method with our CG representation, it is not possible to calculate the helical axis using the local coordinate frame as done with atomistic representations.[62] The helical axis is approximated using the method presented in Ref. 63. Tetrads of nucleotides separated by 3 nucleotides on the sense and anti-sense strands are used to generate a local helical axis. This local helical axis provides a good approximation of the true helix axis, and is used here to determine the major and minor groove widths, the base rise, and to calculate the persistence length of dsDNA.

B. Persistence length

A number of strategies are available for calculation of the persistence length.[64] One definition for flexible chains is

$$l_p = \frac{\langle R_e^2 \rangle}{2L}, \tag{20}$$

where R_e is the end-to-end distance, L is the contour length, and the angle brackets represent a long-time average. For more rigid polymers, a useful definition comes from the bond autocorrelation function,

$$\langle \mathbf{u}(0) \cdot \mathbf{u}(s) \rangle = e^{-s/l_p}. \tag{21}$$

In Eq. (21) the position along the polymer contour length is characterized by s. The vectors $\mathbf{u}(s)$ are tangent to the helical axis at s, calculated using the tetrads described in Sec. III A.

Experimental measurements of ssDNA persistence length range from 2–4 nm (6–12 bp).[65] For long DNA oligomers, comprising 100 bases or more, many persistence lengths of ssDNA are simulated and Eq. (20) is an appropriate choice. The contour length L is calculated using the same ssDNA charge spacing (4.3 Å) that was used in Sec. II E. The persistence length of dsDNA (40–60 nm; 100–150 bp[66]) is much larger than that of ssDNA, thus making Eq. (21) more appropriate.

C. Melting temperatures

The melting temperature of dsDNA provides an important measure of the thermal stability of our DNA model. In previous work,[9,11,32] melting temperatures were determined via a simple parallel tempering or replica exchange simulation (REMD)[67,68] approach. That approach relied on a specific definition of a base pair which depends on the value adopted for the pair cutoff distance. Changing this cutoff by a small amount, however, can shift the position of the melting point.[14] It has also been pointed out that base sites in prior versions of the 3SPN model can form more than one base pair at a time.[28,29]

To circumvent these issues, in this work we employ metadynamics simulations[48,69–71] to estimate the free energy surface associated with hybridization and hairpin formation. In order to efficiently explore phase space, a biasing potential is accumulated as the simulation proceeds. This pushes the system away from local free energy minima and forces the exploration of other regions of phase space. Upon convergence, the added biasing potential can be related to the underlying free energy surface.[71] The following discussion is framed in the context of determining the melting temperature of a duplex; however, the same approach is applicable for calculation of hairpin melting temperatures. Subtle differences are discussed in the supplementary material.[49]

We employ a two-dimensional order parameter that describes the spatial separation and rotational orientation of the two DNA strands. The first order parameter measures the spatial separation between two strands. The distance between the centers of the two interacting single strands is denoted by δ_C. The sites used to describe the center of an ssDNA are restricted to the central five bases of each single strand in the case of oligonucleotides with an odd number of bases, and the central six bases otherwise. The second order parameter is θ_{DNA}, and describes the angle between vectors pointing from the 5′ to the 3′ termini of each single strand. This two-dimensional order parameter surface is able to distinguish between dehybridized (large δ_C) and hybridized DNA (small δ_C and $\theta_{DNA} \approx 180^\circ$ as dsDNA adopts an antiparallel configuration).

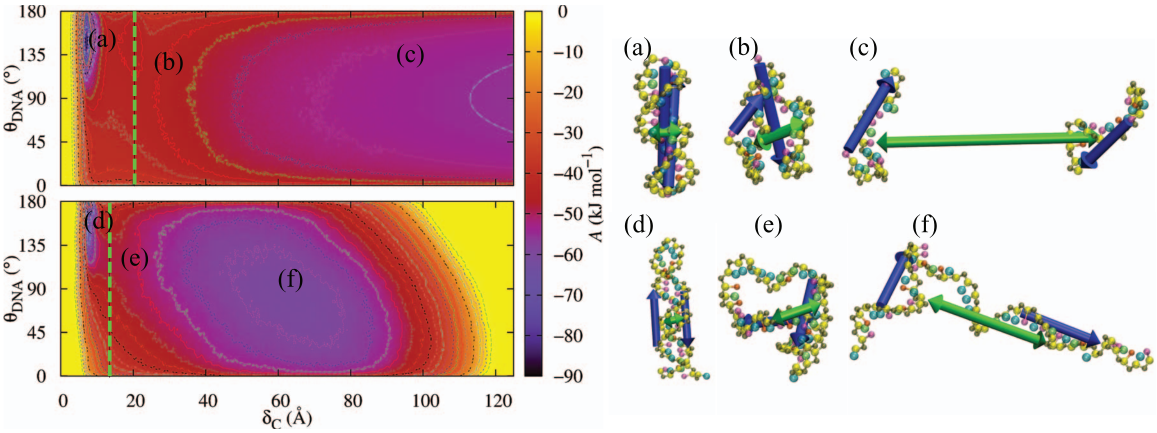

FIG. 4. Metadynamics, combined with an appropriate choice of order parameter, is used to quantify melting temperatures with 3SPN.2. The top panel shows a representative free energy surface for a DNA duplex (5′-TACTAACATTAACTA-3′) at the melting temperature calculated using the order parameters described in the text. (a)–(c) The positions in order parameter space of the configurations found on the right. The dashed green line indicates the approximate position of the saddle point used to differentiate between hybridized and dehybridized states. Note that beyond 90 Å the free energy surface has been extended analytically. The bottom panel shows a representative free energy surface for a DNA hairpin (5′-ATGCAATGCTACATATTCGCTTTTTGCGAATATGTAGCATTGCAT-3′) at its melting temperature using the same order parameters as in the case of the duplex. (d)–(f) The locations of the configurations shown on the right. Again, the dashed green line on the free energy surface is used to separate hybridized and dehybridized states.

Mathematically, the order parameter δ_C is given by

$$\delta_C = \left\| \frac{\sum_{n_B} m_i \boldsymbol{r}_i}{\sum_{n_B} m_i} - \frac{\sum_{n_A} m_j \boldsymbol{r}_j}{\sum_{n_A} m_j} \right\|, \tag{22}$$

where m_i is the mass of site i, $\boldsymbol{r}_i$ is the position in Cartesian space of site i, and n_A and n_B are the indices of the specific coarse-grained sites used to characterize δ_C between single strands A and B, respectively. The order parameter θ_{DNA} is given by

$$\theta_{DNA} = \arccos\left(\frac{\boldsymbol{v}_A \cdot \boldsymbol{v}_B}{\|\boldsymbol{v}_A\|\|\boldsymbol{v}_B\|}\right), \tag{23}$$

where $\boldsymbol{v}_i$ is a vector that points from the COM of the 5′ terminal nucleotide to the COM of the 3′ terminal nucleotide of DNA single strand i (these vectors are depicted as blue arrows in Fig. 4).

Each production run was preceded by 50 ns of equilibration in the NVT ensemble at the appropriate simulation temperature with all walkers but one starting from a dehybridized configuration. Conventional metadynamics simulations were performed until the free energy surface (FES) was found to fluctuate around a steady value. The uncertainty in the predicted free energy surface was calculated by using instantaneous snapshots of the biasing potential from the last half of the simulation. For δ_C, the width of the Gaussian biasing potential was 0.25 Å while the width for θ_{DNA} was 2°. Gaussians with a height of 0.1 kJ mol^{-1} were deposited every 500 time steps for each of the walkers employed for all metadynamics simulations[72] in order to accelerate sampling.

An artificial wall was applied to δ_C to prevent spending significant simulation time exploring dehybridized configurations where entropy dominates. This repulsive wall confines δ_C to separations less than a threshold distance. Beyond this distance, the free energy surface can be extended analytically to any center-to-center separation. Details regarding this extension of the free energy surface are provided in the supplementary material.[49]

In order to sample the melting transition, simulations were performed for temperatures in the vicinity of the melting temperature of an oligonucleotide. For each temperature, the probability of residing in the hybridization basin was calculated. The demarcation between hybridized and dehybridized states on the free energy surface was taken as the saddle point separating the hybridized DNA basin and the dehybridized basin as shown in Fig. 4. As configurations near the bottom of these basins dominate each state's partition function, the exact location of this boundary is unimportant; shifting it by a few angstroms does not affect the calculated probabilities.

To account for concentration effects, the free energy surface was extended to a center-to-center separation that corresponds to the concentration of the experiment against which we are comparing. This extension is not performed for hairpins because strand concentration does not affect hairpin melting temperatures. We also incorporate the correction proposed by Ouldridge and co-workers[34] to account for fluctuations in concentration that cannot be captured in a simulation with only two single strands. This correction is given by

$$f_\infty = \left(1 + \frac{1}{2\Phi}\right) - \sqrt{\left(1 + \frac{1}{2\Phi}\right)^2 - 1}, \tag{24}$$

where f_∞ is the corrected estimate of the fraction of hybridized DNA molecules in the bulk and Φ is the ratio of the probabilities of hybridized and dehybridized states in the simulation, $\Phi = \frac{P_{\text{hybridized}}}{P_{\text{dehybridized}}}$. Plotting $1 - f_\infty$ versus temperature yields melting curves analogous to those obtained experimentally; the melting temperature is defined here as the temperature at which the hybridized and dehybridized states in the thermodynamic limit are equally probable ($f_\infty = 0.5$).

D. Hybridization rate constants

The hybridization of complementary ssDNA has been studied in considerable detail (see Ref. 73 for a summary of key experimental findings). Based on experimental observations, DNA hybridization has been proposed to occur via nucleation involving a few consecutive, in-register W–C base pairs, followed by rapid cooperative zippering. The original hybridization experiments were performed with relatively long sequences of DNA and were performed at high ionic strength. Less is known regarding the hybridization of short DNA oligomers at low ionic strength.[74,75]

The present model is well-suited for examining the mechanisms and rates of such systems because of the base–level resolution provided by the 3SPN representation. We calculate hybridization rate constants using Forward Flux Sampling (FFS);[76] specifically, we use the Rosenbluth FFS algorithm. FFS requires that n interfaces be defined according to an order parameter λ_i that separates the initial and final states. It provides both transition paths between these two states as well as the associated rate constants k expressed as

$$k = \Phi_0 \prod_{i=0}^{n-1} P(\lambda_{i+1}|\lambda_i), \tag{25}$$

where Φ_0 is the flux of trajectories through the initial interface λ_0 and P is the probability of crossing interface λ_{i+1} subject to having crossed λ_i in the direction of λ_n.

The prefactor Φ_0 is calculated here using the methods developed by Northrup *et al.*[77] They separate the association reaction into centrosymmetric and non-centrosymmetric regimes, and probabilistic correction factors κ_1, κ_2, and κ_3 are used to account for finite size effects. The κ's are defined in terms of three cut-off distances d_1, d_2, and d_3. The definitions of the cut-off distances and the calculated probabilities can be found in the supplementary material.[49] The bimolecular reaction rate constant can then be expressed as

$$k = \frac{k_D(d_2)\left[\frac{\kappa_2}{1-(1-\kappa_2)\kappa_3}\right]\alpha}{1-(1-\alpha)\left\{\kappa_1 + \left[\frac{\kappa_2}{1-(1-\kappa_2)\kappa_3}\right](1-\kappa_1)\right\}}, \tag{26}$$

where the $k_D(d_2) = 4\pi D_0 d_2$ is the Smoluchowski result for spherical particles with isotropic reactivity for a COM separation $r = d_2$, D_0 is the relative diffusion constant and α is the overall conditional probability calculated using FFS,

$$\alpha = \prod_{i=0}^{n-1} P(\lambda_{i+1}|\lambda_i). \tag{27}$$

The overall conditional probability α is equivalent to $P(\lambda_n|\lambda_0)$, where λ_0 corresponds to the $\delta_{\mathrm{COM}} = d_1$ and no base pairs formed.

For the present analysis the interfaces λ_i are defined as a linear function of δ_{COM} and the number of W–C base pairs between the complementary sequences. This is shown schematically in Fig. 5 where χ is the normalized COM separation, Γ is the fraction of total possible W–C base pairs formed, and λ_i is depicted by the diagonal lines. Doing so creates two well defined states: dehybridized DNA, with no base pairs being formed and a large δ_{COM}, and fully hybridized DNA with all possible native base pairs being formed and a small δ_{COM}. Note that by including δ_{COM}, we avoid the problems that arise when using only the number of W–C base pairs to differentiate states in parallel tempering calculations. In the present analysis, a base pair is defined as a base pairing interaction of at least 80% of the maximum possible energy and the number of interfaces n was set equal to the total number of bases of each sequence.

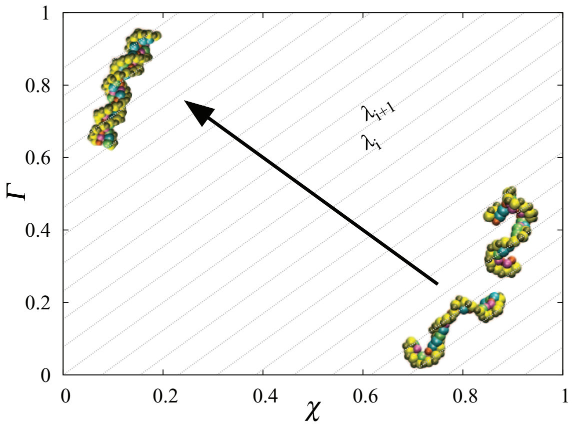

FIG. 5. Interfaces used in FFS to calculate the reaction probability. χ is the normalized COM separation, Γ is the fraction of total possible W–C base pairs formed, and λ_i is depicted by the diagonal lines.

IV. RESULTS

A. Structural properties

Structural properties were calculated using a 32 bp sequence at ionic strength of $I = 100$ mM and at $T = 293.15$ K, simulated for 1 μs with snapshots taken every 2000 time steps. For reference, the same sequence was also simulated using 3SPN.1. The mean and standard deviations of all structural properties are provided in Table I, along with the corresponding experimental quantities. We find good agreement with experimental base rise, helix width, and the number of bases per turn. Table I also reveals that the major and minor grooves of 3SPN.2 are stable through the entire simulation, and are consistent with the values reported by Stofer and Lavery.[61] The existence of stable grooves is important for capturing some aspects of DNA-protein interactions.

TABLE I. Comparison of structural properties predicted by 3SPN.1 and 3SPN.2 to values from the B-DNA crystal structure. Uncertainties represent one standard deviation. Experimental values taken from Refs. 54 and 61. Structural properties were obtained from the 32 bp sequence 5′-ATACAAAGGTGCGAGGTTTCTATGCTCCCACG-3′ at $I = 100$ mM and $T = 293.15$ K.

	3SPN.1	3SPN.2	Expt.
Base rise (Å)	3.26 ± 0.14	3.35 ± 0.04	3.4
Helix width (Å)	23.1 ± 0.7	22.9 ± 0.1	23.0
Bases per turn	10.6 ± 0.3	10.1 ± 0.1	10.0
Major groove (Å)	13.8 ± 1.7	16.6 ± 1.2	17.1
Minor groove (Å)	15.4 ± 1.3	11.7 ± 1.3	11.8

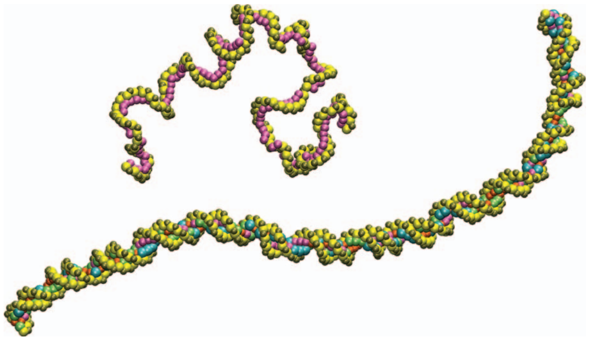

FIG. 6. Representative 144 bp configurations of ssDNA (top) and dsDNA (bottom) from separate 3SPN.2 simulations at $T = 300$ K and $I = 150$ mM.

B. Persistence length

As mentioned earlier, previous versions of the 3SPN model were able to accurately capture the persistence length of dsDNA[11] or ssDNA,[9] but not both. Sequences from λ-phage DNA after digest using the *Taq*I restriction enzyme were simulated at $T = 300$ K and varying ionic strength. Persistence lengths were calculated using the methods outlined in Sec. III B. Representative configurations of ssDNA and dsDNA from simulations are shown in Fig. 6.

As shown in Table II, the persistence lengths of dsDNA sequences are between 40 nm and 60 nm, consistent with the experimental values.[65] The calculated ssDNA persistence lengths for λ-phage DNA sequences were biased as they tended to form secondary structures such as hairpins. We performed persistence length calculations for ssDNA poly(A) sequences in order to obtain results undistorted by hairpin structures. The calculated ssDNA persistence lengths are 2–4 nm, also in agreement with experimental values.[78] The persistence length of ssDNA increases with increasing contour length. This is likely a consequence of our inability to precisely calculate the contour length of ssDNA.

The 3SPN.2 model includes explicit electrostatics, and it is therefore of interest to examine the dependence of the persistence length of dsDNA on ionic strength. The ionic strength directly determines the flexibility of dsDNA via the degree of shielding of electrostatic repulsion between phosphate sites. Nonlinear Poisson–Boltzmann (P–B) theory for uniformly charged cylinders predicts that the persistence length can be expressed as the sum of two contributions:[65] a non-electrostatic portion l_{p_0} and an electrostatic portion l_{el} that is a function of the Debye λ_D and Bjerrum lengths λ_B, i.e.,

$$l_p = l_{p_0} + l_{el} = l_{p_0} + \frac{\lambda_D^2}{4\lambda_B}, \tag{28}$$

with λ_B and λ_D as defined previously. Other coarse-grained models with explicit electrostatics have sought to capture this trend. These models tend to overpredict the dependence of persistence length on ionic strength.[9,20] An exception is provided by the work of Savelyev and Papoian,[79] who presented a 1-site-per-nucleotide model with explicit ions that qualitatively captures this dependence on ionic strength. We have not included explicit ions in 3SPN.2 but, as explained earlier, we have adjusted the effective charge of the phosphate to implicitly account for counter-ion condensation. Figure 7 shows that 3SPN.2, with the adjusted effective charge, gives reasonable agreement with P–B theory. Note that some deviations do occur, most noticeably below $I = 10$ mM. These deviations could be due to the fact that the inter-charge spacing from ssDNA was used to assign the effective charges for the model.

TABLE II. Comparison of persistence length as a function of polymer length. ssDNA simulations were performed with poly(A) sequences to ensure that hairpin formation did not bias persistence length estimates. The quantity f_{CG} denotes the fraction of CG content in each sequence. 5′ base and 3′ base give the location of the sequence within the λ-phage genome. Error bars represent one standard deviation. $T = 300$ K, $I = 150$ mM.

N_{nt}	f_{CG}	5′ base	3′ base	l_p^{ss} (nm)	l_p^{ds} (nm)
68	0.529	43826	43893	…	48.6 ± 1.3
144	0.472	33157	31406	…	48.5 ± 2.4
250	0.528	577	720	…	50.8 ± 5.7
68	0.0	…	…	2.0 ± 0.1	…
144	0.0	…	…	2.5 ± 0.4	…
250	0.0	…	…	3.7 ± 0.8	…

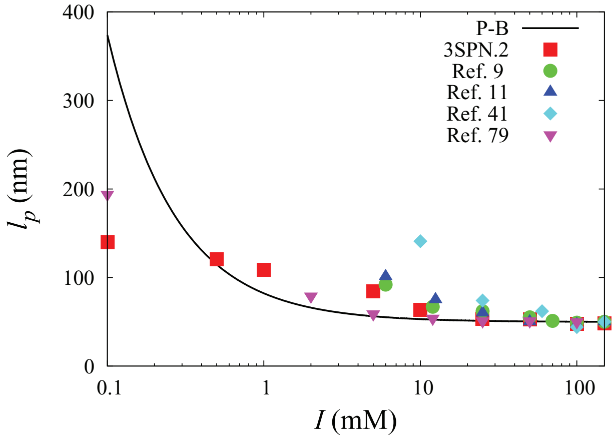

FIG. 7. Scaling of persistence length (l_p) with ionic strength I for various DNA coarse-grained models with explicit charges. All results have been scaled such that the persistence length at $I = 150$ mM is 500 Å.

C. Melting temperatures

Melting temperatures were calculated for both DNA duplexes and hairpins. With 3SPN.2, hairpins can also be simulated because we can now account for the true flexibility of ssDNA. After adjusting for the effect of concentration on the melting temperature, a sigmoidal function of the form

$$g(T) = \frac{1}{1 + e^{A(T-T_m)}} \tag{29}$$

is fit to the values of probability of being in the hybridized state, f_∞, as evaluated with metadynamics calculations at different temperatures.

In Eq. (29) A represents the width of the transition and T_m is the melting temperature. The melting temperature predicted from simulation and the corresponding experimental value are shown in Fig. 8. The average difference between them is of a few Kelvin (see the supplementary material[49] for the

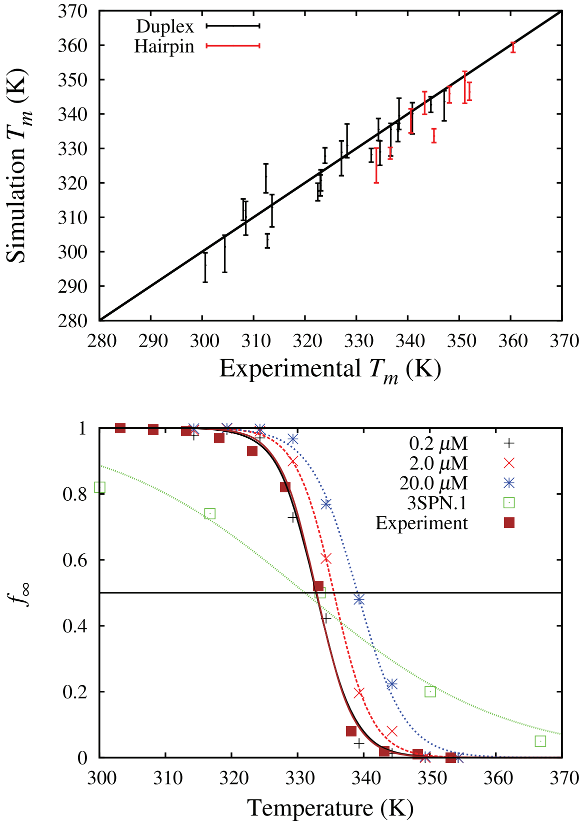

FIG. 8. (Top) Agreement between experimental and simulated melting temperature for duplexes and hairpins. (Bottom) Melting curves obtained from simulation showing the effect of oligomer concentration on calculated melting temperatures. The 4–5 K change in the measured melting temperature is consistent with experiment. The experimental data presented are from an oligomer concentration of 2 μM.[46] Sequence: 5′-TACTTCCAGTGCTCAGCGTA-3′; $I = 69$ mM.

agreement between experiment and simulation presented in tabular form). The width of the ssDNA–dsDNA melting transition is also shown in Fig. 8. In general, the agreement between simulated melting curves and experiment is satisfactory. Here we note again that the dependence on oligomer concentration is captured by extending the free energy surface, as explained in Sec. III C and the supplementary material.[49] The width of the hybridization transition is narrower than that predicted by 3SPN.1, and comparable to that observed in prior models using angle-dependent potentials.[17,18] While a previous report[18] concluded that non-W–C base pairing was the source of the reduction in broadness of the transitions, we observe a reduction due only to the inclusion of anisotropic inter-strand base–base interactions. As the complementary strands melt, base–base interactions can be disrupted not only by further separation but also by rotation to unfavorable angles. This leads to more abrupt and more realistic melting as thermal fluctuations begin to destabilize the double helix.

D. Rate constants

Hybridization rate constants for 4 different sequences at ionic strengths of 100 mM and 500 mM were calculated using FFS as explained in Sec. III D. These systems were consistent with those studied in the work of Gao *et al.*[75] Figure 9 illustrates the correspondence between simulation and experimental values.

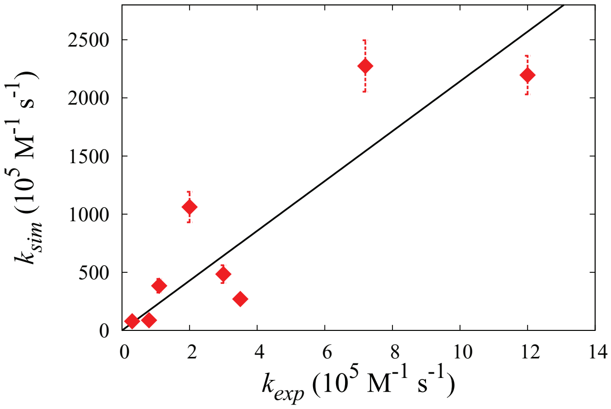

FIG. 9. Comparison of simulated hybridization rate constants to experimental rates constants from Ref. 75. Error bars represent the standard error in the mean.

The simulated rate constants are one to two orders of magnitude larger than experiment. In general, faster dynamics are to be expected from a model where the underlying free energy surface has been coarsened considerably, thus lowering kinetic barriers. For 3SPN.2, we find that agreement with experiment is better at near-physiological ionic strengths (100 mM) than at high ionic strengths (500 mM). Overall, however, the relative magnitude of the predicted rate constants is qualitatively consistent with that observed in experiments. We view this agreement as satisfactory, particularly given the model's simplicity and the fact that no dynamic data were used in its parameterization, besides the friction coefficient.

V. DISCUSSION

A. Improvements

The changes to previous versions of the 3SPN model presented here address several limitations of past models. In what follows, we summarize the main improvements:

- Stable minor and major grooves: 3SPN.2 exhibits stable major and minor grooves. This is now possible because the Gō-like potentials have been supplanted by angle-dependent potentials. While previous versions of the 3SPN model were theoretically capable of resolving the grooves, these were difficult to discern in simulations. Now that the major and minor grooves are properly resolved, 3SPN.2 can be used to study DNA-protein binding. Additionally, the geometry of the grooves may have a subtle influence on the mechanism of hybridization.
- Molecular flexibility: 3SPN.2 captures the correct flexibility of ssDNA and dsDNA. Indeed, the primary motivation for developing 3SPN.2 was to correct an overly large rigidity of ssDNA in an earlier model that was suspected of adversely affecting the mechanism

of DNA hybridization. Both the angle-dependent potentials, as well as the Gaussian well used to model the dihedral potentials, allow the single strand to rotate around the phosphate-sugar backbone. These interactions also cooperate upon helix formation to provide the correct rigidity for dsDNA. By capturing the correct flexibilities of ssDNA and dsDNA, 3SPN.2 can be used to examine biophysical processes involving denaturation and hybridization with better fidelity. This has been demonstrated by calculating rate constants in Sec. IV D.

- Electrostatic interactions: the effective charge of each phosphate site has been reduced in magnitude from -1.0 to -0.6. This modification removed the need for the sugar-sugar interactions invoked in earlier models. Such a potential was primarily needed to facilitate hybridization at low ionic strength ($I < 200$ mM).

In light of the aforementioned improvements, it is worthwhile to revisit previous results obtained with earlier versions of the model. We limit ourselves to applications concerned with hybridization and exclude studies of melting[29] or extensions of the model.[30–32]

Simulations of hybridization using 3SPN.1 resorted to transition path sampling (TPS),[23,25] extended ensemble density of states (EXEDOS),[24] and umbrella sampling.[26] Using each of these techniques, it was observed that hybridization occurred via zippering for heterogeneous sequences or "slithering" for homogeneous sequences. The nucleation event, or formation of initial contacts that form between strands, was often observed to occur with the two strands offset (or out of register) by 2–3 base pairs. The same mechanism was observed in simulations of surface hybridization.[25–27]

The offset nucleation observed in past work was influenced by the relatively rigid, helical structure of ssDNA in 3SPN.1. Because the two strands were predominantly helical, the only way for hybridization to succeed was for the strands to wrap around each other, followed by "slithering." While slithering may not be entirely unphysical, we do believe that it was overemphasized in previous versions of 3SPN.1. In a system consisting of complementary, homogeneous sequences, for example, "slithering" could be observed via the propagation of base pairs, facilitated by thermal fluctuations. In these systems, no large energy barriers would be present to deter the sliding motion of strands.

Using 3SPN.2, we observe that zippering of complementary strands is the dominant mechanism of hybridization. Offset configurations do occur; however, slithering is no longer the primary mechanism for correcting the mismatch.

B. Possible applications

The 3SPN.2 model presented here should be appropriate for studies of DNA hybridization. It is well suited for studies examining the effect of ionic strength on the mechanism of hybridization. It is also easily extensible to studies involving explicit ions. To this end, 3SPN.2 has been implemented within the LAMMPS MD package[80] and is available as a user package (USER-3SPN2). Work is on-going to extend the LAMMPS implementation of 3SPN.2 to include explicit ions, as done previously for 3SPN.1.[32]

The new model is also well-suited to studies of DNA-protein interactions. The resolution of 3SPN.2 is consistent with existing coarse-grained models of proteins[81] and the coupling between the two is currently being considered. In addition, because the topology of dsDNA simulated by 3SPN.2 is derived from a crystal structure, it is possible to create sequence-dependent topologies with minimal adjustments to the parameters presented here. This can be done by adjusting the equilibrium distances and angles for the bonded and non-bonded interactions. All other parameters (K, ϵ's) are unchanged. Such adjustments, which will be discussed in a future publication in the context of nucleosome positioning, allow for configurations with sequence-dependent major and minor groove widths and intrinsic curvature. A similar model could find widespread use for examination of the origin of binding affinities of proteins for particular sequences of DNA.

C. Limitations

The 3SPN.2 model invokes several approximations that merit further discussion. First, electrostatics are treated at the level of a Debye-Hückel approximation. This is strictly appropriate only for low ionic strengths. At high ionic strengths, however, electrostatic interactions are screened and one recovers a "neutral" version of the model analogous to that adopted in other coarse grained representations of DNA. While 3SPN.2 captures the correct melting temperatures at both low and high ionic strengths, it is unclear how appropriate the model is for calculations of dynamic properties at moderate to high ionic strength.

Excluded volume is represented by an isotropic potential. Representing planar base sites as spheres represents a drastic approximation. However, as excluded volume interactions are only experienced by base sites participating in cross-stacking interactions or mismatches, this simplification does not appear to have serious consequences. Importantly, it is not necessary to parameterize entirely new interactions when coupling this model with a coarse-grained protein model. More realistic representations of the base, such as those of Refs. 19 and 21, could serve the twofold purpose of more accurately representing the excluded volume and potentially remove the need for angle-dependent potentials.

The topology of 3SPN.2 is built around the crystal structure of B-DNA. This is of little importance when simulating dsDNA, as the structure fluctuates around the equilibrium structure. However, the model may not be appropriate for studying severely deformed DNA. This includes globally deformed structures such as S-DNA, as well as local deformations induced by the intercalation of dyes (e.g., YOYO[82]) between the base sites.

The choice of reference structure may also have significant repercussions for ssDNA. As ssDNA behaves as a random coil, it has no reference structure. However, we have used the B-DNA as a first approximation for the reference structure of ssDNA. The residual structure in the backbone required for the double helix could limit the possible configurations for

TABLE III. Cartesian and cylindrical polar coordinates for the 3SPN.2 representation of DNA. The masses and the excluded volume diameters of each site are also included. The molecular topology of a single strand is built from the 3′ end using a transformation directly related to the helical rise (3.38 Å) and twist (36°) of B-DNA. For example, if a sugar site is placed at (r, ϕ, and z), the next sugar site moving in the 3′ to 5′ direction will be placed at (r, $\phi + 36°$, and $z + 3.38$ Å. The sites of the complementary strands are related by a dyad along the x-axis; for a sugar site at (x, y, z), the sugar site of the complementary nucleotide will be located at x, $-y$, $-z$. For base sites, the values of r, ϕ, x, and y, and z that are used should correspond to the identity of the site being placed. For additional details, see Ref. 45.

Site type	x (Å)	y (Å)	z (Å)	r (Å)	ϕ (°)	m (amu)	σ (Å)
Phosphate (P)	−0.628	8.896	2.186	8.918	94.035	94.97	4.5
Sugar (S)	2.365	6.568	1.280	6.981	70.196	83.11	6.4
Adenine (A)	0.296	2.489	0.204	2.506	83.207	134.1	5.4
Thymine (T)	−0.198	3.412	0.272	3.418	93.327	125.1	7.1
Guanine (G)	0.542	2.232	0.186	2.297	76.349	150.1	4.9
Cytosine (C)	0.683	3.265	0.264	3.336	78.192	110.1	6.4

ssDNA. The consequences of this approximation are not yet understood.

VI. CONCLUSION

We have presented a new **3**-**S**ite-**P**er-**N**ucleotide coarse-grained model (3SPN.2) that has been extensively validated using enhanced sampling techniques such as metadynamics and forward flux sampling. The model provides good agreement with experimental measures of structural properties such as duplex width, base rise, and major and minor groove width. In addition, it captures the persistence length of both ss- and dsDNA and predicts melting temperatures that are consistent with experiment. The 3SPN.2 model is also shown to predict hybridization rate constants that qualitatively agree with experimental values.

The new 3SPN.2 model could be useful in several areas of computational biophysics and materials science. These include studies of the mechanisms of nucleic acid hybridization and DNA-protein binding and the structural properties of nano-engineered DNA-hybrid materials, DNA origami, and DNA liquid crystals. It has been implemented in the LAMMPS MD package and is available for download.

ACKNOWLEDGMENTS

The authors thank Dr. Gregory N. Toepperwein for helpful discussions. The University of Wisconsin–Madison Center for High Throughput Computing is gratefully acknowledged for providing resources and computer expertise. D.M.H., G.S.F., and J.K.W. were funded in part through Training Grant T32HG002760 to the Genomic Sciences Training Program. D.M.H. was also supported by a Graduate Research Fellowship from the National Science Foundation (NSF). This work was supported by the NSF through the University of Wisconsin–Madison Nanoscale Science and Engineering Center (Grant No. DMR0832760). We also thank anonymous referees whose comments led to significant improvements to this paper.

APPENDIX A: MODEL TOPOLOGY

Each site in the 3SPN.2 model is placed at the respective center-of-mass of the phosphate, sugar, or base, as calculated from the crystal structure of B-DNA.[45] Isotropic excluded volume interactions are included between all sites at least 2 nucleotides away within the same strand and all sites on other strands with one exception: base sites experience excluded volume interactions only between inter-strand sites that are non-complementary. Complementary sites experience a shorter-range repulsion determined by the equilibrium distance of A-T and G-C base pairs. Table III provides coordinates and masses for each site, as well as the algorithm used to create the configurations used as the reference structure. Table IV gives the equilibrium separations σ_{AT} and σ_{GC} of the W–C base pairs.

APPENDIX B: MODEL PARAMETERS

The force constants for the bonded interactions are listed in Table IV. Also included in the table are the parameters α_i and K_i that modulate the range of the Morse potential and the width of the attractive cones depicted in Fig. 2. Base pairing interactions are longer-ranged (smaller value of α) than cross-stacking interactions. Base stacking is shorter-ranged than base pairing because the fundamental basis for these interactions is π–π stacking, instead of hydrogen bonding.

TABLE IV. Table of 3SPN.2 force field parameters used in the bonded and non-bonded interactions.

Parameter	Value
k_b	0.6 kJ/mol/Å^2
k_θ	200 kJ/mol/rad^2
k_ϕ	6.0 kJ/mol
ϵ_{r}	1.0 kJ/mol
K_{BS}	6.0
α_{BS}	3.0
K_{CS}	8.0
α_{CS}	4.0
K_{BP}	12.0
α_{BP}	2.0
σ_{AT}	5.941 Å
σ_{GC}	5.518 Å
ϵ_{AT}	16.73 kJ/mol
ϵ_{GC}	21.18 kJ/mol

TABLE V. Equilibrium bond lengths r_o, bend angles θ_o, and dihedral angles ϕ_o. The direction of the bonds is important. P(5′) or S(5′) represents the phosphate or sugar, respectively, in the 5′ direction of the adjacent site while P(3′) or S(3′) represents the phosphate or sugar in the 3′ direction.

Bond	r_o (Å)			Bend	θ_o (°)
P(5′)–S	3.899			S–P–S	94.49
S–P(3′)	3.559			P–S–P	120.15
S–A	4.670			P–S–A	103.53
S–T	4.189			P–S–T	92.06
S–G	4.829			P–S–G	107.40
S–C	3.844			P–S–C	103.79
				A–S–P	112.07
Dihedrals		ϕ_o (°)	σ_ϕ	T–S–P	116.68
(5′)P–S–P–S(3′)		−154.79	0.30	G–S–P	110.12
(5′)S–P–S–P(3′)		−179.17	0.30	C–S–P	110.33

The values of K were parameterized in such a way that behavior such as kinking was discouraged. The relative permissiveness of the modulating function f is justified as follows: Base pairing is the most restrictive (largest K), as it represents directional base pairing. Cross-stacking is less restrictive than base pairing, as no hydrogen bonds are being formed. Base stacking is least restrictive because adjacent bases are free to slide or shift as DNA is deformed.

All short-ranged non-bonded interactions (excluded volume and base-base interactions) were calculated using a cutoff of 18 Å. Electrostatic repulsions were cutoff beyond 50 Å, except in calculations performed below 50 mM ionic strength. At such low ionic strength, the Debye length is large and this long-range cutoff must be extended to a sufficiently long distance ($\sim 4\lambda_D$).

The equilibrium bond lengths, bend angles, and dihedral angles are provided in Table V. These equilibrium angles were measured from the coarse-grained topology that can be generated using the coordinates found in Table III.

The equilibrium distances, angles, and strengths used in the non-bonded base-base interactions are shown in Tables VI and VII. The angles and distances are obtained from the coarse-grained topology, just as was done for the bonded parameters. The strength of the base-base interactions deserves additional discussion.

The base stacking energies were parameterized by performing simulations of systems analogous to those of Protozanova *et al.*[44] In their work, they determined the free energies of stacking by measuring the electrophoretic mobility of nicked double helices. We simulated nicked double helices using two collective variables in metadynamics: the distance between the stacking bases and the angle between the two helical segments flanking the nick. Additional details can be found in the supplementary material.[49] Using Boltzmann inversion,[47] we iteratively adjusted the values of the stacking strength according to the following relationship

$$\epsilon_{i+1} = \epsilon_i + kT \ln\left(\frac{F_i}{F_{exp}}\right), \tag{B1}$$

where ϵ is the stacking strength between 2 bases, i is the iteration number, F_i is the calculated free energy using ϵ_i, and F_{exp} is the experimental free energy. Repeated Boltzmann inversion cycles were performed until satisfactory agreement was achieved.

The strengths of the base pairing and cross-stacking interactions were parameterized using the Santalucia nearest-neighbor enthalpies.[43] The nearest-neighbor enthalpies were divided into base pairing and cross-stacking energies. The ratio of the G–C to A–T Watson–Crick base pairing strength was 1.266, as done in the previous version of 3SPN.[11] The cross-stacking interactions were constrained to be 12% of the total base step energies, as observed in Ref. 83. The base step energies were reduced by the average W–C base pairing energy between the two bases in the base step and their complements. The remaining energy was then divided equally into 2 separate cross-stacking interactions. An adjustable parameter was used to scale uniformly the nearest-neighbor enthalpies until agreement was achieved between simulation and experimental melting temperatures for sequence B at $I = 69$ mM (see the supplementary material[49] for exact sequence). Using this value of the adjustable parameter, melting temperatures were predicted for other sequences, including hairpins at varying ionic strengths.

TABLE VI. Reference angles used to modulate U_{bp} and U_{cstk}. The indices i and j correspond to the identity of the base sites being used to define the vector r_{ij}. All angles are expressed in degrees.

		Base j								
		ϕ_{1o}					θ_{1o}			
		A	T	G	C		A	T	G	C
	A	…	−38.35	…	…	A	…	156.54	…	…
	T	−38.35	…	…	…	T	135.78	…	…	…
	G	…	…	…	−45.81	G	…	…	…	154.62
	C	…	…	−45.81	…	C	…	…	152.74	…
Base i		θ_{2o}					θ_{3o}			
		A	T	G	C		A	T	G	C
	A	…	135.78	…	…	A	…	116.09	…	…
	T	156.54	…	…	…	T	116.09	…	…	…
	G	…	…	…	152.74	G	…	…	…	131.78
	C	…	…	154.62	…	C	…	…	131.78	…

TABLE VII. Values of the strengths ϵ_{ij}, equilibrium distances σ_{ij}, and equilibrium angles θ_{XX_o} for the base stacking (a) and cross-stacking interactions (b-c). The arrows ↑ and ↓ represent the sense and anti-sense strands, respectively with the bases participating in the base pair indicated by $^{5'}\uparrow$ and $\downarrow^{3'}$. $_{3'}\uparrow$ and $\downarrow_{5'}$ indicate adjacent bases in the 3′ and 5′ directions, respectively, that participate in cross-stacking interactions.

(a)								Base $_{3'}\uparrow$							
			ϵ (kJ/mol)					σ (Å)					θ_{BS_o} (°)		
Base $^{5'}\uparrow$		A	T	G	C		A	T	G	C		A	T	G	C
	A	14.39	14.34	13.25	14.51	A	3.716	3.675	3.827	3.975	A	101.15	85.94	105.26	90.26
	T	10.37	13.36	10.34	12.89	T	4.238	3.984	4.416	4.468	T	101.59	89.50	104.31	90.82
	G	14.81	15.57	14.93	15.39	G	3.576	3.598	3.664	3.822	G	100.89	84.83	105.48	90.18
	C	11.42	12.79	10.52	13.24	C	3.859	3.586	4.030	3.957	C	115.95	101.51	119.32	104.49

(b)								Base $\downarrow_{5'}$							
			ϵ (kJ/mol)					σ (Å)					θ_{CS_o} (°)		
Base $^{5'}\uparrow$		A	T	G	C		A	T	G	C		A	T	G	C
	A	2.186	2.774	2.833	1.951	A	6.208	6.876	6.072	6.941	A	154.38	159.10	152.46	157.58
	T	2.774	2.186	2.539	2.980	T	6.876	7.480	6.771	7.640	T	147.10	153.79	144.44	148.59
	G	2.833	2.539	3.774	1.129	G	6.072	6.771	5.921	6.792	G	154.69	157.83	153.43	158.60
	C	1.951	2.980	1.129	4.802	C	6.941	7.640	6.792	7.698	C	160.37	164.45	158.62	162.73

(c)								Base $_{3'}\uparrow$							
			ϵ (kJ/mol)					σ (Å)					θ_{CS_o} (°)		
Base $\downarrow^{3'}$		A	T	G	C		A	T	G	C		A	T	G	C
	A	2.186	2.774	2.980	2.539	A	5.435	6.295	5.183	5.965	A	116.88	121.74	114.23	114.58
	T	2.774	2.186	1.951	2.833	T	6.295	7.195	6.028	6.868	T	109.42	112.95	107.32	106.41
	G	2.980	1.951	4.802	1.129	G	5.183	6.028	4.934	5.684	G	119.34	124.72	116.51	117.49
	C	2.539	2.833	1.129	3.774	C	5.965	6.868	5.684	6.453	C	122.10	125.80	120.00	119.67

The cross-stacking energies are specified in terms of the two bases between which the cross-stacking interaction is occurring. This is done to easily relate thermodynamic base step enthalpy to cross-stacking interactions where the strands of ssDNA come together out-of-register. In that situation, using the cross-stacking energies derived from the enthalpy of a single base step is not appropriate; even though a W–C base pair has formed, the bases adjacent to the W–C base pair may not be complementary. Therefore, the cross-stacking strength between base types i and j on the sense and anti-sense strands, respectively, is generally different from the cross-stacking strength between base types j and i on the sense and anti-sense strands. The cross-stacking strengths between the sense W–C–forming base and an adjacent base site on the anti-sense strand and the cross-stacking strengths between the anti-sense W–C-forming base and an adjacent base site on the sense strand are found in Table VII.

We remind the reader that cross-stacking interactions are only experienced by bases participating in a W–C base pair and adjacent bases on opposite strands. Thus, for an AG base step interacting with its complement (CT), one cross-stacking interaction is defined between the A site on the sense strand and the C site on the anti-sense strand. A second cross-stacking interaction is defined between the G site and the T site. If the base site on the sense strand is participating in a W–C pair with a base at the 5′ end of the anti-sense strand, then it does not participate in cross-stacking. Likewise, a base site on the anti-sense strand does not participate in a cross-stacking interaction if its W–C complement is at the 3′ end of the sense strand.

[1] A. Perez, F. J. Luque, and M. Orozco, Acc. Chem. Res. **45**, 196 (2012).

[2] A. P. Lyubartsev and L. Nordenskiöld, J. Phys. Chem. B **101**, 4335 (1997).

[3] A. A. Kornyshev and S. Leikin, Proc. Natl. Acad. Sci. U.S.A. **95**, 13579 (1998).

[4] H. Jian, A. V. Vologodskii, and T. Schlick, J. Comput. Phys. **136**, 168 (1997).

[5] R. M. Jendrejack, J. J. de Pablo, and M. D. Graham, J. Chem. Phys. **116**, 7752 (2002).

[6] P. T. Underhill and P. S. Doyle, J. Non-Newton. Fluid **122**, 3 (2004).

[7] K. Drukker and G. C. Schatz, J. Phys. Chem. B **104**, 6108 (2000).

[8] H. L. Tepper and G. A. Voth, J. Chem. Phys. **122**, 124906 (2005).

[9] T. A. Knotts, N. Rathore, D. C. Schwartz, and J. J. de Pablo, J. Chem. Phys. **126**, 084901 (2007).

[10] P. Poulain, A. Saladin, B. Hartmann, and C. Prévost, J. Comput. Chem. **29**, 2582 (2008).

[11] E. J. Sambriski, D. C. Schwartz, and J. J. de Pablo, Biophys. J. **96**, 1675 (2009).

[12] M. Maciejczyk, A. Spasic, A. Liwo, and H. A. Scheraga, J. Comput. Chem. **31**, 1644 (2009).

[13] S. Niewieczerza and M. Cieplak, J. Phys.: Condens. Matter. **21**, 474221 (2009).

[14] P. D. Dans, A. Zeida, M. R. Machado, and S. Pantano, J. Chem. Theory Comput. **6**, 1711 (2010).

[15] S. M. Gopal, S. Mukherjee, Y.-M. Cheng, and M. Feig, Proteins **78**, 1266 (2010).

[16] A. Morriss-Andrews, J. Rottler, and S. S. Plotkin, J. Chem. Phys. **132**, 035105 (2010).

[17] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Chem. Phys. **134**, 085101 (2011).

[18] M. C. Linak, R. Tourdot, and K. D. Dorfman, J. Phys. Chem. **135**, 205102 (2011).

[19] A. V. Savin, M. A. Mazo, I. P. Kikot, L. I. Manevitch, and A. V. Onufriev, Phys. Rev. B **83**, 245406 (2011).

[20] C. W. Hsu, M. Fyta, G. Lakatos, S. Melchionna, and E. Kaxiras, J. Chem. Phys. **137**, 105102 (2012).

[21] L. E. Edens, J. A. Brozik, and D. J. Keller, J. Phys. Chem. B **116**, 14735 (2012).

[22] Y. He, M. Maciejczyk, S. Ołdziej, H. A. Scheraga, and A. Liwo, Phys. Rev. Lett. **110**, 098101 (2013).

[23] E. J. Sambriski, D. C. Schwartz, and J. J. de Pablo, Proc. Natl. Acad. Sci. U.S.A. **106**, 18125 (2009).

[24] E. J. Sambriski, V. Ortiz, and J. J. de Pablo, J. Phys.: Condens. Matter **21**, 034105 (2009).

[25] M. J. Hoefert, E. J. Sambriski, and J. J. de Pablo, Soft Matter **7**, 560 (2011).

[26] T. J. Schmitt and T. A. Knotts IV, J. Chem. Phys. **134**, 205105 (2011).

[27] T. J. Schmitt, J. B. Rogers, and T. A. Knotts IV, J. Chem. Phys. **138**, 035102 (2013).

[28] V. Ortiz and J. J. de Pablo, Phys. Rev. Lett. **106**, 238107 (2011).

[29] A.-M. Florescu and M. Joyeux, J. Chem. Phys. **135**, 085105 (2011).

[30] T. R. Prytkova, I. Eryazici, B. Stepp, S.-B. Nguyen, and G. C. Schatz, J. Phys. Chem. B **114**, 2627 (2010).

[31] R. C. DeMille, T. E. Cheatham, and V. Molinero, J. Phys. Chem. B **115**, 132 (2011).

[32] G. S. Freeman, D. M. Hinckley, and J. J. de Pablo, J. Chem. Phys. **135**, 165104 (2011).

[33] T. X. Hoang and M. Cieplak, J. Chem. Phys. **113**, 8319 (2000).

[34] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Phys.: Condens. Matter **22**, 104102 (2010).

[35] M. R. Machado, P. D. Dans, and S. Pantano, Phys. Chem. Chem. Phys. **13**, 18134 (2011).

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

[37] F. Romano, A. Hudson, J. P. K. Doye, T. E. Ouldridge, and A. A. Louis, J. Chem. Phys. **136**, 215102 (2012).

[38] C. Matek, T. E. Ouldridge, A. Levy, J. P. K. Doye, and A. A. Louis, J. Phys. Chem. B **116**, 11616 (2012).

[39] P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. K. Doye, and A. A. Louis, J. Chem. Phys. **137**, 135101 (2012).

[40] M. C. Linak, R. Tourdot, and K. D. Dorfman, J. Phys. Chem. **137**, 205102 (2012).

[41] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).

[42] A. V. Savin, I. P. Kikot, M. A. Mazo, and A. V. Onufriev, Proc. Natl. Acad. Sci. U.S.A. **110**, 2816 (2013).

[43] J. SantaLucia, Proc. Natl. Acad. Sci. U.S.A. **95**, 1460 (1998).

[44] P. Protozanova, E. Yakovchuk, and M. D. Frank-Kamenetskii, J. Mol. Biol. **342**, 775 (2004).

[45] S. Arnott, P. J. Campbell Smith, and R. Chandrasekaran, *Atomic Coordinates and Molecular Conformations for DNA-DNA, RNA-RNA, and DNA-RNA Helices*, Vol. 2 of CRC Handbook of Biochemistry and Molecular Biology, 3rd ed. (CRC Press, Cleveland, 1976), pp. 411–422.

[46] R. Owczarzy, Y. You, B. G. Moreira, J. A. Manthey, L. Huang, M. A. Behlke, and J. A. Walder, Biochemistry **43**, 3537 (2004).

[47] D. Reith, M. Pütz, and F. Müller-Plathe, J. Comput. Chem. **24**, 1624 (2003).

[48] A. Laio and F. L. Gervasio, Rep. Prog. Phys. **71**, 126601 (2008).

[49] See supplementary material at http://dx.doi.org/10.1063/1.4822042 for details regarding model parameterization and free energy and rate calculations.

[50] W. Humphrey and A. Dalke, J. Mol. Graphics **14**, 33 (1996).

[51] F. Oosawa and M. Kasai, J. Mol. Biol. **4**, 10 (1962).

[52] G. Manning, J. Chem. Phys. **51**, 924 (1969).

[53] W. K. Olson and G. S. Manning, Biopolymers **15**, 2391 (1976).

[54] V. A. Bloomfield, D. M. Crothers, and I. Tinoco, Jr., *Nucleic Acids: Structure, Properties, and Functions* (University Science Books, 2000).

[55] S. Liu and M. Muthukumar, J. Chem. Phys. **116**, 9975 (2002).

[56] A. Stogryn, IEEE T. Microw. Theory **19**, 733 (1971).

[57] A. Catenaccio, Y. Daruich, and C. Magallanes, Chem. Phys. Lett. **367**, 669 (2003).

[58] D. P. Fernandez, A. R. H. Goodwin, E. W. Lemmon, J. M. H. Levelt Sengers, and R. C. Williams, J. Phys. Chem. Ref. Data **26**, 1125 (1997).

[59] G. Bussi and M. Parrinello, Phys. Rev. E **75**, 056707 (2007).

[60] A. E. Nkodo, G. M. Garnier, B. Tinland, C. Desruisseaux, L. C. McCormick, G. Drouin, and G. W. Slater, Electrophoresis **22**, 2424 (2001).

[61] E. Stofer and R. Lavery, Biopolymers **34**, 337 (1994).

[62] R. Lavery and H. Sklenar, J. Biomol. Struct. Dyn. **6**, 655 (1989).

[63] P. C. Kahn, Comput. Chem. **13**, 185 (1989).

[64] P. C. Hiemenz and T. P. Lodge, *Polymer Chemistry*, 2nd ed. (CRC Press, 2007).

[65] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **94**, 6185 (1997).

[66] S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).

[67] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).

[68] Q. Yan and J. J. de Pablo, J. Chem. Phys. **111**, 9509 (1999).

[69] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 12562 (2002).

[70] S. Singh, C.-c. Chiu, and J. J. de Pablo, J. Stat. Phys. **145**, 932 (2011).

[71] S. Singh, M. Chopra, and J. J. de Pablo, Annu. Rev. Chem. Biomol. Eng. **3**, 369 (2012).

[72] P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello, J. Phys. Chem. B **110**, 3533 (2006).

[73] J. G. Wetmur, Annu. Rev. Biophys. Bioeng. **5**, 337 (1976).

[74] M. E. Craig, D. M. Crothers, and P. Doty, J. Mol. Biol. **62**, 383 (1971).

[75] Y. Gao, L. K. Wolf, and R. M. Georgiadis, Nucleic Acids Res. **34**, 3370 (2006).

[76] R. J. Allen, C. Valeriani, and P. Rein ten Wolde, J. Phys.: Condens. Matter **21**, 463102 (2009).

[77] S. H. Northrup, S. A. Allison, and J. A. McCammon, J. Chem. Phys. **80**, 1517 (1984).

[78] B. Tinland, A. Pluen, J. Sturm, and G. Weill, Macromolecules **30**, 5763 (1997).

[79] A. Savelyev and G. A. Papoian, Proc. Natl. Acad. Sci. U.S.A. **107**, 20340 (2010).

[80] See http://lammps.sandia.gov for LAMMPS MD package.

[81] H. Kenzaki, N. Koga, N. Hori, R. Kanada, W. Li, K.-i. Okazaki, X.-Q. Yao, and S. Takada, J. Chem. Theory Comput. **7**, 1979 (2011).

[82] K. Günther, M. Mertig, and R. Seidel, Nucleic Acids Res. **38**, 6526 (2010).

[83] C. A. Johnson, R. J. Bloomingdale, V. E. Ponnusamy, C. A. Tillinghast, B. M. Znosko, and M. Lewis, J. Phys. Chem. B **115**, 9244 (2011).