THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Thermodynamics of DNA Hybridization from Atomistic Simulations

Published as part of The Journal of Physical Chemistry virtual special issue "Emily A. Carter Festschrift".

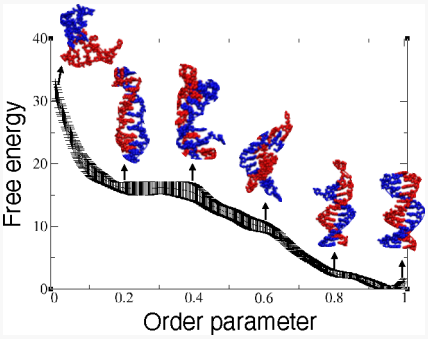
Gül H. Zerze, Frank H. Stillinger, and Pablo G. Debenedetti*

Cite This: *J. Phys. Chem. B* 2021, 125, 771−779

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Studying DNA hybridization equilibrium at atomistic length scales, either via molecular dynamics (MD) or through commonly used advanced sampling approaches, is notoriously difficult. In this work, we describe an order-parameter-based advanced sampling technique to calculate the free energy of hybridization, and estimate the melting temperature of DNA oligomers at atomistic resolution. The free energy landscapes are reported as a function of a native-topology-based order parameter for the Drew−Dickerson dodecamer and for a range of DNA decamer sequences of different GC content. Our estimated melting temperatures match the experimental numbers within ±15 °C. As a test of the numerical reliability of the procedures employed, it was verified that the predicted free energy surfaces and melting temperatures of the D- and L-enantiomers of the Drew−Dickerson dodecamer were indistinguishable within numerical accuracy.



INTRODUCTION

DNA hybridization, which involves noncovalent attachment of two strands of DNA chains to one another, resulting in a double-stranded helix (duplex), is a fundamental biophysical process. The double-stranded helix may unwind back to the coil-like single-stranded configurations via, for example, thermal melting, as repeatedly happens during a polymerase chain reaction. The thermodynamics of this process has been extensively studied in the laboratory, especially via thermal experiments, which have formed the foundations of elegant theories of nucleic acid melting, such as the nearest-neighbor or next-nearest-neighbor models.[1−7] These experimental studies have also helped the simulation community to develop coarse-grained representations of nucleic acids for studying hybridization thermodynamics.[8−15]

Coarse-grained models of nucleic acids typically incorporate base-stacking, base-pairing, and some backbone dihedral interactions. These models have been successful in capturing the thermodynamics and melting temperatures of DNA chains up to lengths of 20 base pairs.[11,12] Atomistic models of nucleic acids, on the other hand, are more commonly used to study fine structural details, configurational variation, and the flexibility of double-stranded helices as well as transitions between different forms of duplexes (i.e., the structural polymorphism in double-stranded DNA).[16−21]

There are comparatively few studies tackling the thermodynamics of DNA hybridization at the atomistic scale, as the rare and slow nature of DNA hybridization poses significant computational challenges. Atomistic models have been used to study the formation of single base pairs in double-stranded DNA[22] and base-stacking free energy in a short oligomer containing two nucleobases.[23] However, studying longer chains (such as decamers) has been out of reach until recently. Piana has presented the first atomistic molecular dynamics (MD) study of helix−coil transitions in short DNA sequences (hexamers)[24] using advanced-sampling MD. Specifically, the author used bias-exchange metadynamics, with pairwise coordination between atoms involved in Watson−Crick pairs and phosphate backbone dihedral angles as the order parameters to advance the sampling. However, the expected stability of the DNA hybrid was not reproduced.

Lomzov et al.[25] used atomistic models to calculate DNA's melting temperature from the difference in the total energy of hybridized states and that of unhybridized states, assuming that (i) the volume change upon hybridization is negligible, (ii) the difference in total energy between double- and single-stranded states is the same as the hybridization enthalpy (which is, of course, correct when assumption i is true), and (iii) hybridization entropy can be derived from a correlation between hybridization enthalpy and entropy. In another work, by Ren and co-workers,[26] the relative hybridization free energy between RNA and modified nucleic acids was estimated using an atomistic polarizable force field via an alchemical transition method. However, we note that, although the calculations in both of these works can yield melting temperatures and

Received: October 11, 2020
Revised: December 20, 2020
Published: January 12, 2021

https://dx.doi.org/10.1021/acs.jpcb.0c09237
J. Phys. Chem. B 2021, 125, 771−779

thermodynamic quantities associated with hybridization, they cannot yield the free energy landscape of DNA hybridization, i.e., the free energy as a function of one or more order parameters.

In this work, we propose a novel advanced-sampling approach to study DNA hybridization thermodynamics using a classical, fully atomistic model. We employ a parallel tempering scheme[27] in conjunction with metadynamics[28−30] using an order parameter inspired by protein folding and binding studies.[31] We calculated the free energy surface and melting temperature of the Drew−Dickerson dodacamer (DDD)[32−34] and found that our melting temperature estimates agree well with the predictions from melting experiments[35] and those from a nearest-neighbor model.[5,6,36] We further tested our proposed technique by simulating decamers of various GC contents and found that our melting temperature estimates match the experimental numbers within ±15 °C. We note that the predictive capability of our simulations is limited by the accuracy of the underlying potential energy function (CHARMM36 nucleic acids). Although this force field is commonly used in atomistic simulations, it is not specifically optimized for predicting the hybridization free energy. Performing such an optimization, however, would not be possible without a proper sampling technique. Therefore, our sampling technique is also a useful asset for future refinements of nucleic acid force fields.

METHODS

DNA Systems and Modeling. We studied the hybridization of a dodecamer and seven decamers. We first showed the validity of our approach using the DDD DNA, also known as *Eco*RI dodecamer.[32−34] To further validate our method, we also tested a number of decamers of systematically varied GC content, for which extensive experimental data are available[37] (sequences from 5′ to 3′ are ATCAATCATA, TTGTAGTCAT, CCAACTTCTT, ATCGTCTGGA, CGATCTGCGA, GATGCGCTCG, and GGGACCGCCT). Finally, we studied a DDD variant, fully composed of L-nucleotides, i.e., the mirror image of the original DDD, in order to measure the impact of a total chiral inversion.

We set up two independent simulations of DDD, where the only difference is in the initial condition. All replicas of one of the simulations were initiated from a completely unhybridized (but otherwise equilibrated) configuration where the two strands are far apart from each other, whereas in the other simulation all replicas were initiated from the hybridized crystal structure (PDB ID: 1BNA). Initial coordinates for decamers were generated using the CHARMM program and the internal coordinate (IC) table supplied in the CHARMM36 force field.[38−40] The chirally inverted DDD was generated by modifying the IC table of the CHARMM36 force field for the CHARMM program (modified IC parameters are available upon request).

Each DNA was solvated in a truncated octahedron box with 5.7 and 6 nm spaced faces for decamers and DDD, respectively. The number of water molecules was adjusted to achieve approximately 1000 kg/m^3 water density. The simulation box was large enough not to allow chain A (or chain B) to interact with itself from any periodic image of the simulation box within the limits of nonbonded interaction cutoff distance.

DDD DNA was studied at 100 mM NaCl concentration, whereas decamers were studied at 69 mM NaCl concentration, following the corresponding experimental protocols.[35,37] Salt concentration was adjusted by adding Na^+ and Cl^- ions to the solution after enough counterions were added to neutralize the net charge of the given DNAs. All DNA systems were modeled using the CHARMM36 nucleic acid force field[39,40] in combination with TIP3P water,[41,42] as this is one of the most widely used force fields in the biomolecular simulation community. Cooordinates and connectivity information are then converted to a format compatible with the GROMACS MD engine.[43]

Simulation Methods. We used a combination of parallel-tempering in the well-tempered ensemble and well-tempered metadynamics (PTWTE-WTM).[27−30,44,45] For the (well-tempered) metadynamics component of this combined advanced sampling technique, we used a similarity-based order parameter to describe the similarity between a given configuration and a chosen reference state. The reference state was chosen as the native hybridized state. In an ideal metadynamics sampling, the order parameter(s) used in order to accelarate the sampling should capture all slow degrees of freedom of a given system. The order parameter that we used, the fraction of native-like contacts between strands, Q_{inter}, accounts for interstrand interactions such as base-pairing and cross-stacking interactions between strands. For other degrees of freedom that may not be directly captured by Q_{inter}, such as intramolecular interactions, e.g., backbone torsional rotations, etc., we combined the (well-tempered) metadynamics with the parallel-tempering (in the well-tempered ensemble) scheme.

The PTWTE[27,30] part of PTWTE-WTM simulations was set up to cover the melting temperature of each oligomer. Therefore, the chosen temperature ranges were between 300 and 475 K for the dodecamer and two of the decamers (f(G·C) = 0.7, 0.8), whereas the range was between 279.5 and 442.5 K for the other decamers studied. In the framework of the well-tempered ensemble, the potential energy was biased using 500 kJ/mol Gaussian width and 1.0 kJ/mol initial Gaussian height, with Gaussian potentials added every 2000 steps and with a bias factor of 16. The temperatures of the 14 replicas were distributed geometrically for all oligomers. The average replica exchange acceptance ratio was 25% for decamers and 20% for the dodecamer.

The WTM part of the simulations used only one order parameter, Q_{inter}. A detailed description of the parameter is presented in the following section. The initial Gaussian height was set to 1.8 kJ/mol with a bias factor of 35 for the WTM sampling on Q_{inter}. The Gaussian width was set to 0.005. Since Q_{inter} is defined strictly between 0 and 1, interval limits were applied together with restraining potentials to avoid accumulating systematic errors at the boundaries of Q_{inter}, restraining the sampling between the Q_{inter} = 0.01 and 0.99 boundaries. The force contribution from metadynamics bias acting on Q_{inter} was set to zero beyond these limits, and harmonic restraining potentials with a spring constant of 75,000 kJ/mol were applied at the defined boundary.[46]

After initial solvation, all DNA systems were equilibrated during 100 ps NVT simulations (T = 300 K) followed by 100 ps NPT simulations (T = 300 K, P = 1 bar). Prior to starting the PTWTE-WTM simulations, unbiased NVT simulations of each replica were performed for 200 ps in order to equilibrate the potential energy of the replicas. Production simulations were run at NPT conditions, where the temperature was maintained constant at the given replica temperature using a

https://dx.doi.org/10.1021/acs.jpcb.0c09237
J. Phys. Chem. B 2021, 125, 771−779

Nosé–Hoover thermostat[47,48] with a 1 ps time constant. Atmospheric pressure (1 bar) was maintained using an isotropic Parrinello–Rahman barostat[49,50] with a time constant of 2 ps. Electrostatic interactions were calculated using the particle-mesh Ewald method[51] with a real space cutoff distance of 1 nm. A cutoff distance of 1 nm was also used for the van der Waals interactions.

All simulations were performed using the GROMACS code (version 2016.3)[52,53] with the PLUMED (version 2.3.1) patch[54,55] for metadynamics sampling.

Description of the Order Parameters. Inspired by protein folding studies, where the folding reaction is described on the native-like contacts coordinate,[31,56−59] we described the DNA hybridization reaction based on the interstrand native-like atomic contacts. We defined the order parameter Q_{inter} as $N_{nb}^{-1}\sum_{(i,j)}\frac{1}{1+\exp(\gamma(r_{ij}-\lambda r_{ij}^0))}$ by adapting the generalized definition of Q,[56] for interstrand contacts. The sum runs over N_{nb} number of atomic pairs (i, j) that are considered in contact where atom i is from strand A and atom j is from strand B (complementary strand). Any heavy (i.e., non-hydrogen) nucleobase (nb) atom of strand A is considered in contact with a heavy nb atom of strand B if the distance between them is less than 5 Å in the reference native structure. r_{ij}^0 and r_{ij} are the distances between i and j in the reference native structure and in any given instantaneous configuration, respectively. γ in the smoothing function was taken as 50 nm^{-1}, and the adjustable parameter λ was taken as 1.5.[31] The Protein Data Bank (PDB) structure 1BNA[32] was used as the reference native structure for DDD. The reference native structures (double-stranded B-DNA configurations) for decamers were created by the Nucleic Acid Builder (nab) tool of AmberTools[60] which uses X-ray fiber diffraction data.[61]

We also analyzed the Watson–Crick hydrogen bonds and number of stacked bases for DDD. Watson–Crick hydrogen bonds form between nucleotides A and T and G and C. In total, DDD forms 12 Watson–Crick base pairs in its native structure, involving 32 specific donor–acceptor atom pairs; i.e., there are 32 Watson–Crick hydrogen bonds in the native state. A pairwise coordination of specific donor–acceptor atoms (from the list of 32 pairs) within 0.6 nm cutoff distance was considered as the formation criterion for a hydrogen bond. A similar distance-based criterion was also applied to the base-stacking order parameter. A stacked pair is considered as a pair of any adjacent nucleotides (within the same strand) whose base centers of mass are within 0.6 nm of each other. A strand of 12 nucleotides can form 11 stacks of bases. We counted stacked pairs separately for each strand. Therefore, the maximum number of stacked bases is 22.

Two-State Analysis. In this work, we assumed that the hybridization reaction of DNA strands follows a two-state equilibrium

$$A + B \leftrightarrow AB \tag{1}$$

where AB represents double-stranded helix (duplex) and A and B are unhybridized, complementary single strands.

The unbiased probability densities of order parameters of interest (including Q_{inter}) were obtained by reweighting all of the biases deposited on the Q_{inter} and potential energy coordinates by using the technique described by Tiwary and Parrinello[62] as implemented in PLUMED (v2.3.1)[55] code. The reweighted probability density, $P(Q_{inter})$, was then converted to free energy as a function of Q_{inter} via the equation

$$F(Q_{inter}) = -kT \ln P(Q_{inter}) + C \tag{2}$$

where kT is the product of the Boltzmann constant and temperature and C is a constant.

We defined the hybridized and unhybridized states on the Q_{inter} coordinate by choosing the basins $0 < Q_{inter} < 0.5$ and $0.5 < Q_{inter} < 1$ as the unhybridized and hybridized states, respectively. Based on this definition, the probabilities (ρ) of unhybridized (A + B) and hybridized (AB) substates were calculated as

$$\rho(A+B) = \frac{\sum_{i=1}^{i=N_{tr}} \exp\frac{-F_i}{kT}}{\sum_{i=1}^{i=N_T} \exp\frac{-F_i}{kT}}, \quad \rho(AB) = \frac{\sum_{i=N_{tr}}^{i=N_T} \exp\frac{-F_i}{kT}}{\sum_{i=1}^{i=N_T} \exp\frac{-F_i}{kT}} \tag{3}$$

at a given temperature T. The index i runs over the number of bins in the discretized Q_{inter} coordinate; i.e., F_i is the free energy corresponding to a given bin on the Q_{inter} coordinate. N_{tr} is the number of bins between $Q_{inter} = 0$ and $Q_{inter} = 0.5$, and N_T is the total number of bins in the entire Q_{inter} space.

For any given description of the hybridized and unhybridized states, the fraction of unhybridized DNA at equilibrium can be calculated as[63]

$$\frac{\rho(A+B)}{\rho(A+B)+\rho(AB)} \tag{4}$$

for each temperature. For our definition of the states, we note that $\rho(A + B) + \rho(AB) = 1$ at any temperature. We define the melting temperature as the temperature at which $\rho(AB) = \rho(A + B) = 0.5$.

To calculate the thermodynamic quantities of hybridization, the probabilities of substates (eq 3) are converted to the relative free energies via

$$G_u = -kT \ln \frac{\sum_{i=1}^{i=N_{tr}} \exp\frac{-F_i}{kT}}{\sum_{i=1}^{i=N_T} \exp\frac{-F_i}{kT}}, \quad G_h = -kT \ln \frac{\sum_{i=N_{tr}}^{i=N_T} \exp\frac{-F_i}{kT}}{\sum_{i=1}^{i=N_T} \exp\frac{-F_i}{kT}} \tag{5}$$

ΔG_{hyb} can then be calculated as the difference between the hybridized and unhybridized free energies:

$$\Delta G_{hyb} = G_h - G_u = -kT\left(\ln \frac{\sum_{i=N_{tr}}^{i=N_T} \exp\frac{-F_i}{kT}}{\sum_{i=1}^{i=N_{tr}} \exp\frac{-F_i}{kT}}\right) \tag{6}$$

The entropy change upon hybridization (ΔS_{hyb}) was calculated from the temperature dependence of ΔG_{hyb}, i.e.,

$$\Delta S_{hyb} = -\left[\frac{\partial \Delta G_{hyb}}{\partial T}\right]_p \tag{7}$$

Then, ΔH_{hyb} was calculated as

$$\Delta H_{hyb} = \Delta G_{hyb} + T\Delta S_{hyb} \tag{8}$$

RESULTS AND DISCUSSION

We first characterized the free energy surface of the DDD's hybridization at 300 K. Sampling of Q_{inter} as a function of time/replica at 300 K can be seen in Figure S1. As the description of the order parameter (Q_{inter}) implies, the configurations with

https://dx.doi.org/10.1021/acs.jpcb.0c09237
J. Phys. Chem. B 2021, 125, 771−779

low Q_{inter} are unhybridized, whereas high Q_{inter} configurations are hybridized (Figure 1). Configurations with Q_{inter} = 1 are identical to the reference (hybridized) state. After forming the initial native-like contacts, the free energy decreases monotonically as Q_{inter} increases, i.e., as more native-like contacts form, consistent with a nucleation-elongation mechanism found in coarse-grained models of DNA.[11,12,64,65]

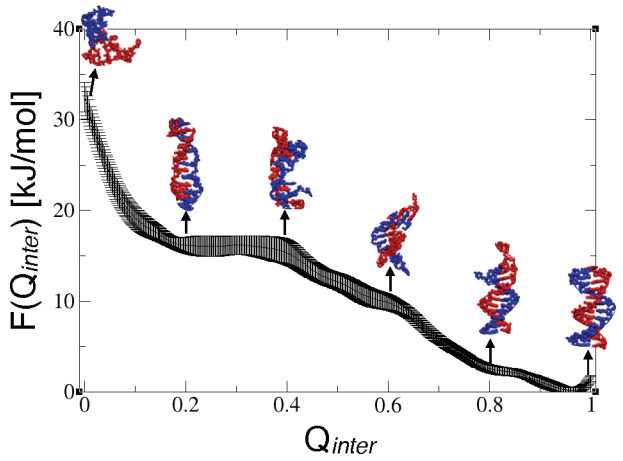


Figure 1. Free energy landscape of the DDD at 300 K. Errors are calculated from the standard deviation of the last 10 free energy profiles, equally separated every 10 ns. Snapshots show representative configurations of the DDD and its complementary strand at selected Q_{inter}.

Thermal Melting of the Drew−Dickerson Dodecamer. We estimated the melting temperatures from the free energies (as a function of Q_{inter}) evaluated at various temperatures (Figure 2A), as described in the Methods section. We chose the unhybridized and hybridized states as the states comprising the order parameter ranges $0 < Q_{inter} < 0.5$ and $0.5 < Q_{inter} < 1$, respectively. The two-state behavior can be visualized clearly on a probability representation (Figure 2B). The melting curve is also reported in Figure 2C, where the experimental and computed melting temperatures are 326.3[35] and 331.8 K, respectively.

To ensure that we attain the sampling equilibrium, we evaluated the unbiased free energy as a function of Q_{inter} at every 10 ns starting at t = 0 and measured ΔG_{hyb} (eq 6) as a function of time (Figure S2). The part of the trajectories where ΔG_{hyb} fluctuates within $\pm kT$ was considered for analysis. As a more stringent criterion of equilibrium, we also compared melting curves from two independent simulations with different initial conditions (see Methods), and we found no dependence on initial conditions (Figure S3).

Although the comparison of our melting curve with the corresponding experimental measurement shows good agreement (Figure 2C), it should be noted that the experimental data from Marky et al.[35] were collected at a strand concentration 3 orders of magnitude smaller than the DNA concentration in our simulation box. DNA concentration measurably affects the melting temperature,[66] as expected from the fact that the physical reaction of DNA hybridization is not equimolar. Strand concentration is further discussed in the next section, where we systematically study the melting temperature of decamers.

The order parameter Q_{inter} is, not surprisingly, correlated with the number of Watson−Crick hydrogen bonds and with the number of stacked bases (Figure 3A and B; see also the Methods section for descriptions). To emphasize the importance of the order parameter selection for the success of sampling, we performed a supporting set of simulations using a similar combined advanced sampling. In this simulation, we used the same PTWTE-WTM setup, except for the order parameter that was biased. Instead of Q_{inter}, we biased the number of Watson−Crick hydrogen bonds. In this sampling scheme, we found that the unhybridized state is always more stable than the hybridized state at 300 K during the entire course of sampling (Figure S4). Therefore, we conclude that the sampling that relies on the Watson−Crick hydrogen bonds cannot accurately predict melting. The two order parameters (Q_{inter} and Watson−Crick hydrogen bonds) being correlated does not necessarily mean that they perform equally well in sampling. Q_{inter} outperforms the Watson−Crick hydrogen bonds, as it provides a more detailed structural description based on native topology. While both order parameters rely on pairwise coordination, Q_{inter} enforces contact formation for each pair within a prescribed tolerance

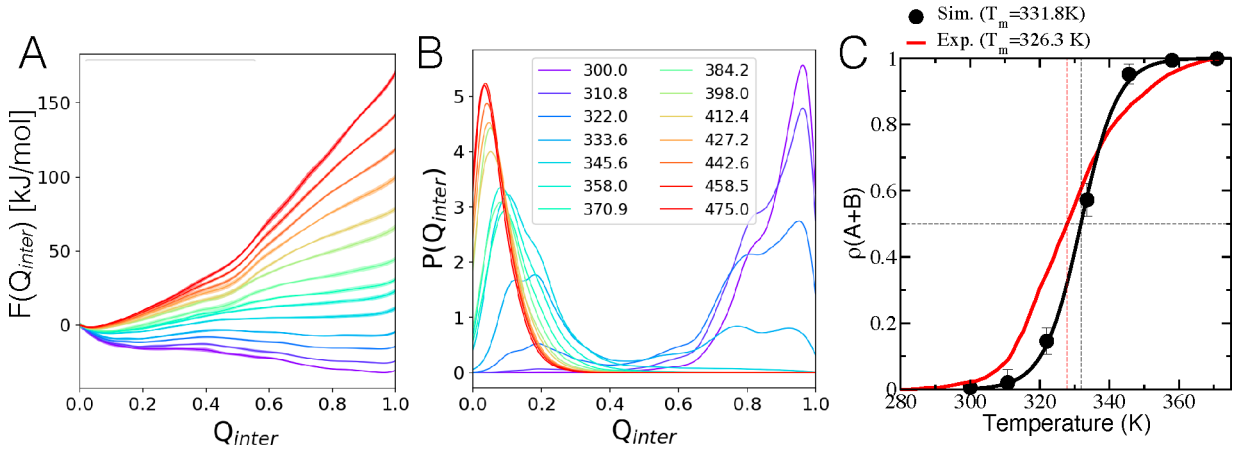


Figure 2. Thermal melting of the DDD. (A) Free energy surfaces of DDD at different temperatures (eq 2). Free energy surfaces are overlaid at Q_{inter} = 0 by adjusting the constant C in eq 2. (B) Probability densities at different temperatures (as shown in the legend in units of K). (C) Melting curve calculated from eq 4. The temperature at which $\rho(A + B)$ = 0.5 is the melting temperature. The experimental data are from thermal melting (UV) experiments.[35]

https://dx.doi.org/10.1021/acs.jpcb.0c09237
J. Phys. Chem. B 2021, 125, 771−779

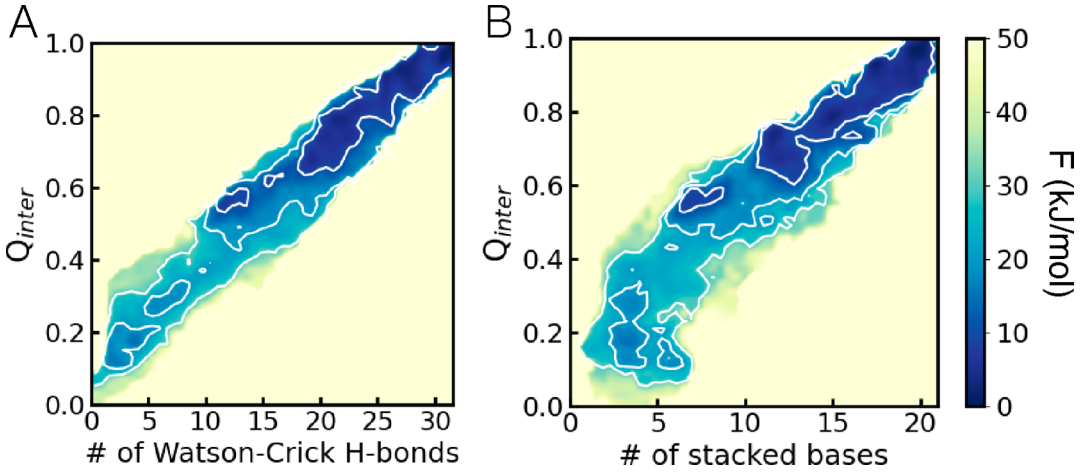


Figure 3. 2D free energy surfaces as a function of Q_{inter} and two other order parameters, number of Watson−Crick hydrogen bonds (A) and number of stacked bases (B) at 300 K (see also the Methods section). The contour lines correspond to 10, 20, and 30 kJ/mol.

relative to the corresponding pair separation in the native structure, whereas native-like Watson−Crick hydrogen bonds enforce a contact formation within a uniform cutoff distance (see Methods). Moreover, Watson−Crick hydrogen bonds are defined by 32 specific atom pairs (donor−acceptor pairs), whereas Q_{inter} is defined by 626 atom pairs between the strands (see Methods). Therefore, Q_{inter} significantly reduces the degeneracy of states along the path between unhybridized and hybridized states compared to Watson−Crick hydrogen bonds.

We note that the unhybridized basin, i.e., the collection of configurations having a small number of native-like contacts (low Q_{inter}), does not necessarily differentiate the configurations where the two strands are far apart from each other from those where two strands are relatively close to each other but unhybridized. The 2D free energy as a function of Q_{inter} and center of mass distance between two strands (Figure 4) shows that the low Q_{inter} state samples a large range of chain separations (between 0 and 2.8 nm) and the separations larger than 2.8 nm are not sampled. Therefore, we emphasize that the hybridization free energy calculations in this work are related to *melting* of native-like contacts, such as Watson−Crick hydrogen bonds, rather than pulling of strands apart from each other. Additional free energy calculations (umbrella sampling) that we performed show that pulling the two strands fully apart from each other involves 3−5 times larger free energy changes than just melting the native-like contacts at 300 K (Figure S5).

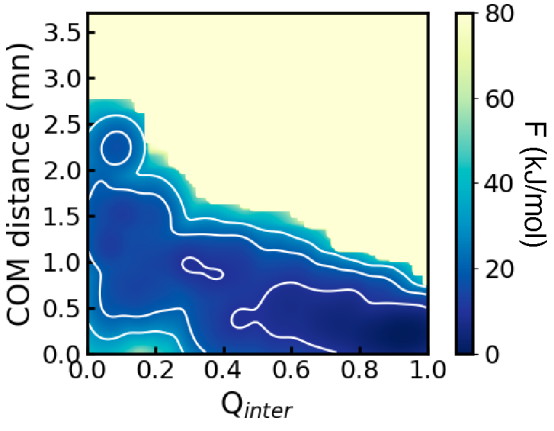


Figure 4. 2D free energy surface of DDD as a function of Q_{inter} and the center of mass (COM) distance between two strands of DDD at 300 K. The contour lines correspond to 10, 20, and 30 kJ/mol. Although large separations of chains (up to 3.5 nm for a given volume of the simulation box) are allowed, they are not sampled.

Correlation between the Experimental and Simulation Melting Temperatures for Varied GC Content. To illustrate the generalized applicability of the method, we performed simulations of DNA oligomers at various GC fractions, f(G·C), at fixed length (decamer) and fixed salt concentration (69 mM NaCl). We chose these oligomers, as extensive experimental data are available for them.[37] For the reference double-stranded helical (hybridized) state, we generate double-stranded helices of the given sequence in B-DNA configuration using the nab tool of AmberTools[60] (see Methods).

Following exactly the same methodology that we described for DDD DNA in the previous section, we calculated the free energy surface of hybridization for various temperatures (Figure 5) and the melting temperatures for each oligomer. A comparison of the melting temperatures calculated from our simulations and the corresponding experimental melting temperatures is presented in Figure 6 for each f(G·C). The maximum difference between simulation and experiment is ±15°.

We note that the experimental melting temperatures are measured at 5 μM strand concentration,[37] whereas our systems are several orders of magnitude more concentrated (30 mM). To keep the system computationally tractable at atomistic resolution, we are limited to concentrations on the order of mM; μM concentrations in all atom, solvent explicit simulations are beyond the limits of commonly available computational power. The melting temperature is expected to decrease upon strand dilution, as there is a larger free volume available that can be occupied by unhybridized strands. Interestingly, however, we obtained the same melting curves within the numerical accuracy upon 2-fold dilution (twice larger volume) for f(G·C) = 0.7 DNA oligomer (Figure S6).

To give a more quantitative idea of the effect of strand concentration on the melting temperature, we note that, for the DDD DNA sequence in 100 mM NaCl solution, the melting temperature predicted by the nearest-neighbor model[5,6,36] increases by approximately 5 °C upon an order of magnitude increase in strand concentration, which is less

https://dx.doi.org/10.1021/acs.jpcb.0c09237
J. Phys. Chem. B 2021, 125, 771−779

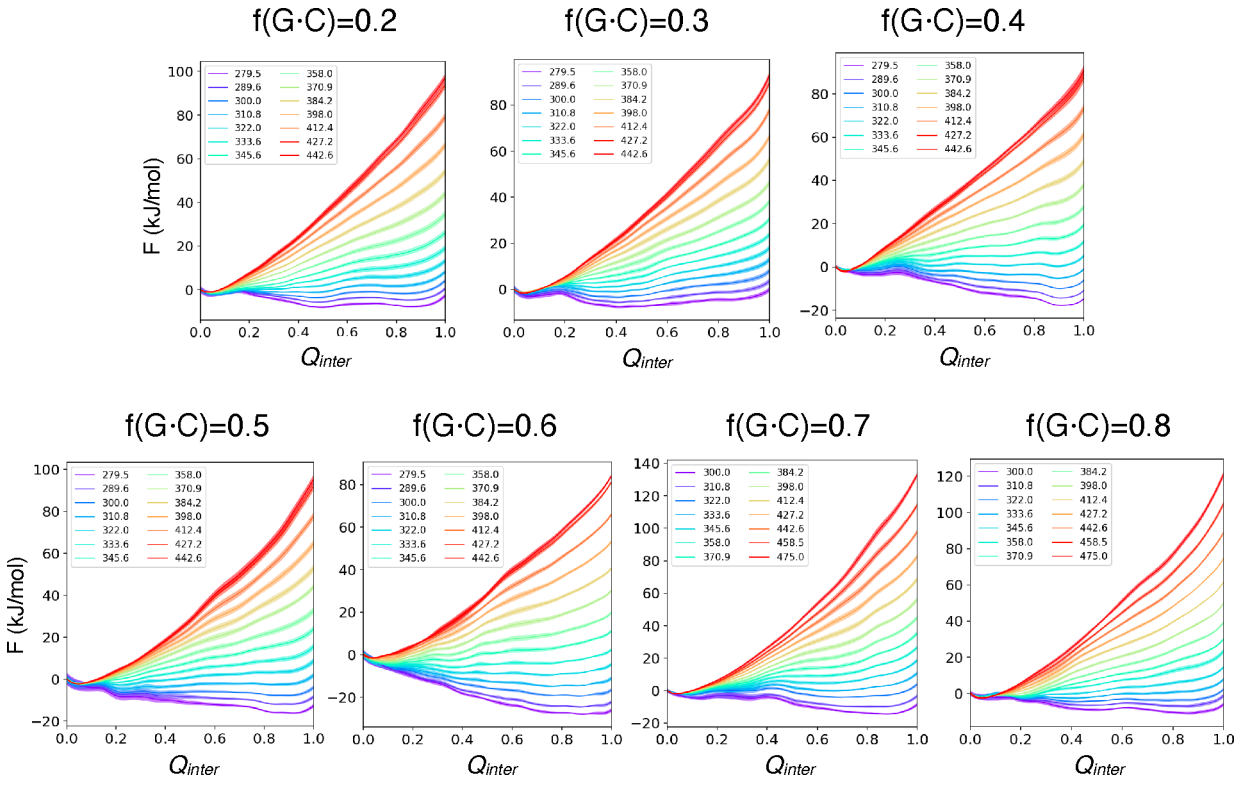


Figure 5. Free energy surface of DNA oligomers of various GC content, evaluated at several temperatures. The shading represents errors estimated from block averaging.

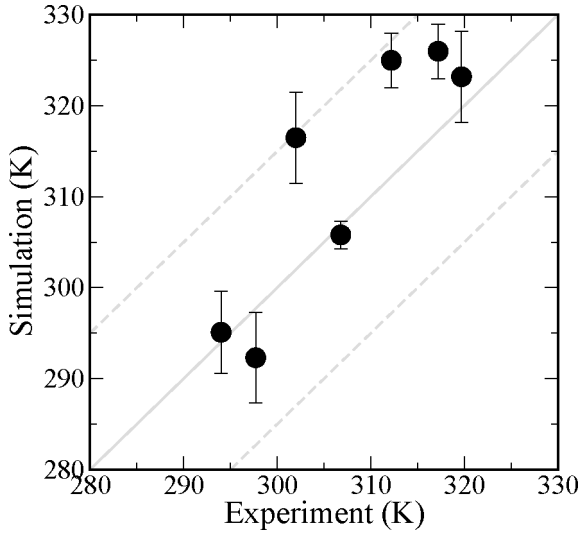


Figure 6. Correlation between the simulation and experimental[37] estimates of melting temperatures. The solid gray line denotes perfect correlation ($y = x$), whereas the broken gray lines correspond to a ±15 °C deviation. The experimental melting temperature increases as f(G·C) increases.

than the maximum deviation (15 °C) between experimental estimates and our atomistic simulation estimates. A systematic investigation of the effect of strand concentration on hybridization thermodynamics and melting temperatures will be the subject of a future investigation.

We also calculated the enthalpic (ΔH_{hyb}) and entropic ($-T\Delta S_{hyb}$) contributions to ΔG_{hyb} (Figure S7, eqs 7 and 8). At low temperatures (below the melting temperature for each DNA), the negative enthalpic contribution dominates, making the hybridized state more stable. At elevated temperatures, the entropic contribution dominates, destabilizing the hybridized state.

Comparison of Hybridization Thermodynamics for D-DNA and L-DNA Chains. Lastly, we confirm that our method is insensitive to chiral inversion. We compared the melting curves of (D-)DDD DNA (naturally occurring) and the corresponding (L-)DDD DNA (Figure 7) under the same salt and strand concentrations and estimated the same melting temperature for both chains (within numerical error). As the order parameter that we used in order to calculate the free energy surface relies solely on intermolecular distances between complementary strands, it is not expected to be affected by mirror symmetry. Accordingly, we used the same reference distances for (L-)DDD DNA and for (D-)DNA.

The chirality of the DNA is imposed in the initial condition, which cannot change during the course of the simulation, as the classical force fields do not allow bond breaking and forming. As there are no asymmetric terms in the nucleic acid force field, we directly transferred all bonded and nonbonded interactions of D-nucleic acids to their L-counterparts. As observed here, the melting curve is not affected by the complete chiral inversion. This is not surprising, as the underlying interaction parameters are exactly the same. The only difference is a corresponding mirror-image switch in the chirality of the resulting double-helix assembly, from right-handed in the case of D-DNA to left-handed for the L-oligomers.

CONCLUSIONS

In this work, we presented an order-parameter-based advanced sampling method that allows access to the free energy landscape and the thermodynamics of DNA hybridization at

https://dx.doi.org/10.1021/acs.jpcb.0c09237
J. Phys. Chem. B 2021, 125, 771−779

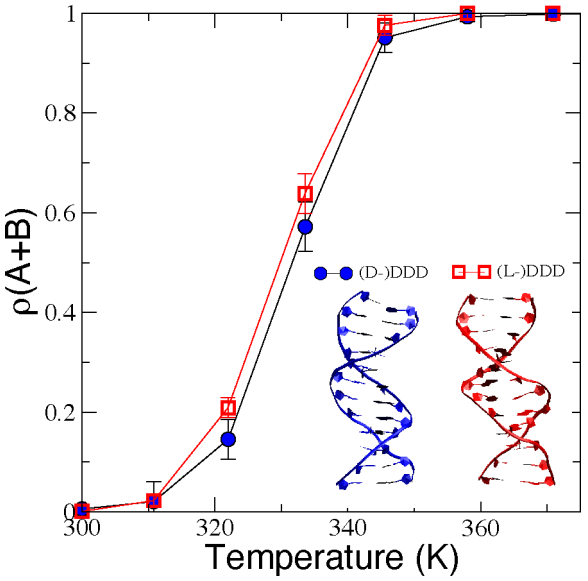


Figure 7. Comparison of melting curves for DDD at its two mirror-image chiral realizations. The blue structure is a representative double-stranded helical configuration of the hybridized state of (D-)DDD (right-handed helix), and the red structure is that of the hybridized state of (L-)DDD, which is a left-handed helix (Q_{inter} = 0.99 for both).

atomistic resolution. This work pushes the limits of atomistic simulations, bypassing the need for a coarse-grained description of the hybridization process that takes place in microseconds to milliseconds for the DNA oligomers.[67] We note that the ±15 °C deviation in melting temperatures with respect to experimental estimates is relatively large compared to the corresponding deviation when using coarse-grained models.[11,12] Our technique satisfies stringest sampling tests: sampling initiated from completely different initial conditions and sampling of the DNA oligomer at two opposite chiral realizations converged to the same melting temperatures and free energy surfaces. Using a native-topology-based order parameter, we were able to obtain a statistically significant number of hybridization transitions and report free energy surfaces at an atomistic scale. However, another important consideration that determines the predictive capability of any simulation is the underlying potential energy function. Atomistic force fields for nucleic acids are not necessarily optimized so as to accurately predict melting temperatures. However, with our technique, such an optimization is also now possible as a future direction.

A particular advantage of our method arises naturally from it being a sampling technique based on metadynamics. It can therefore be coupled with additional order parameters via, for example, concurrent metadynamics[45] or parallel-biasing metadynamics.[68] This coupling would be especially useful when there is a competing reaction of interest for a given nucleic acid system. An example of such a system would include a competition between surface adsorption of nucleic acid strands and hybridization.[69] Studying the equilibrium between these competing reactions would be useful to develop cancer sensing platforms. Further future directions of inquiry include characterization of free energy surfaces of RNA folding in various environments.[70,71]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c09237.

Seven supporting figures showing the sampling of the order parameter as a function of time, change in the free energy as a function of time, melting curves from two independent simulations, free energy profiles from simulations where an alternative order parameter is biased, free energy as a function of center of mass distance, melting curves of different strand concentrations, and enthalpic and entropic contributions to the free energy (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Pablo G. Debenedetti − *Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, United States;* orcid.org/0000-0003-1881-1728; Email: pdebene@princeton.edu

Authors

Gül H. Zerze − *Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, United States;* orcid.org/0000-0002-3074-3521

Frank H. Stillinger − *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States;* orcid.org/0000-0002-1225-8186

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.0c09237

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

G.H.Z. thanks Daniel Kozuch and Dr. Pablo Piaggi for useful discussions. P.G.D. acknowledges the support of the National Science Foundation (grant CHE-1856704). The simulations presented in this work are performed on computational resources managed and supported by Princeton Research Computing, a consortium of groups including the Princeton Institute for Computational Science and Engineering (PICSciE) and the Office of Information Technology's High Performance Computing Center and Visualization Laboratory at Princeton University.

■ REFERENCES

(1) Borer, P. N.; Dengler, B.; Tinoco, I., Jr; Uhlenbeck, O. C. Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.* **1974**, *86*, 843−853.

(2) Gotoh, O.; Tagashira, Y. Stabilities of nearest-neighbor doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles. *Biopolymers* **1981**, *20*, 1033−1042.

(3) Wartell, R. M.; Benight, A. S. Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Phys. Rep.* **1985**, *126*, 67−107.

(4) Aida, M. An ab initio molecular orbital study on the sequence-dependency of DNA conformation: an evaluation of intra-and inter-strand stacking interaction energy. *J. Theor. Biol.* **1988**, *130*, 327−335.

(5) Owczarzy, R.; Vallone, P. M.; Gallo, F. J.; Paner, T. M.; Lane, M. J.; Benight, A. S. Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers* **1997**, *44*, 217−239.

(6) SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 1460−1465.
(7) SantaLucia, J.; Allawi, H. T.; Seneviratne, P. A. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **1996**, *35*, 3555−3562.
(8) Drukker, K.; Wu, G.; Schatz, G. C. Model simulations of DNA denaturation dynamics. *J. Chem. Phys.* **2001**, *114*, 579−590.
(9) Knotts, T. A., IV; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. A coarse grain model for DNA. *J. Chem. Phys.* **2007**, *126*, 084901.
(10) Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics. *J. Chem. Theory Comput.* **2010**, *6*, 1711−1725.
(11) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* **2011**, *134*, 085101.
(12) Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; De Pablo, J. J. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **2013**, *139*, 144903.
(13) Cragnolini, T.; Derreumaux, P.; Pasquali, S. Coarse-grained simulations of RNA and DNA duplexes. *J. Phys. Chem. B* **2013**, *117*, 8047−8060.
(14) Maffeo, C.; Ngo, T. T.; Ha, T.; Aksimentiev, A. A coarse-grained model of unstructured single-stranded DNA derived from atomistic simulation and single-molecule experiment. *J. Chem. Theory Comput.* **2014**, *10*, 2891−2896.
(15) He, Y.; Liwo, A.; Scheraga, H. A. Optimization of a Nucleic Acids united-RESidue 2-Point model (NARES-2P) with a maximum-likelihood approach. *J. Chem. Phys.* **2015**, *143*, 243111.
(16) Cheatham, T., III; Kollma, P. Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* **1996**, *259*, 434−444.
(17) Pérez, A.; Luque, F. J.; Orozco, M. Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.* **2012**, *45*, 196−205.
(18) Dans, P. D.; Danilane, L.; Ivani, I.; Drsata, T.; Lankas, F.; Hospital, A.; Walther, J.; Pujagut, R. I.; Battistini, F.; Gelpi, J. L.; Lavery, R.; Orozco, M.; et al. Long-timescale dynamics of the Drew-Dickerson dodecamer. *Nucleic Acids Res.* **2016**, *44*, 4052−4066.
(19) Dans, P. D.; Balaceanu, A.; Pasi, M.; Patelli, A. S.; Petkeviciute, D.; Walther, J.; Hospital, A.; Bayarri, G.; Lavery, R.; Maddocks, J.; et al. The static and dynamic structural heterogeneities of B-DNA: extending Calladine-Dickerson rules. *Nucleic Acids Res.* **2019**, *47*, 11090−11102.
(20) Balaceanu, A.; Buitrago, D.; Walther, J.; Hospital, A.; Dans, P. D.; Orozco, M. Modulation of the helical properties of DNA: next-to-nearest neighbour effects and beyond. *Nucleic Acids Res.* **2019**, *47*, 4418−4430.
(21) Kuzmanic, A.; Dans, P. D.; Orozco, M. An in-depth look at DNA crystals through the prism of molecular dynamics simulations. *Chem.* **2019**, *5*, 649−663.
(22) Hagan, M. F.; Dinner, A. R.; Chandler, D.; Chakraborty, A. K. Atomistic understanding of kinetic pathways for single base-pair binding and unbinding in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13922−13927.
(23) Norberg, J.; Nilsson, L. Temperature dependence of the stacking propensity of adenylyl-3′, 5′-adenosine. *J. Phys. Chem.* **1995**, *99*, 13056−13058.
(24) Piana, S. Atomistic Simulation of the DNA Helix- Coil Transition. *J. Phys. Chem. A* **2007**, *111*, 12349−12354.
(25) Lomzov, A. A.; Vorobjev, Y. N.; Pyshnyi, D. V. Evaluation of the Gibbs free energy changes and melting temperatures of DNA/DNA duplexes using hybridization enthalpy calculated by molecular dynamics simulation. *J. Phys. Chem. B* **2015**, *119*, 15221−15234.
(26) Jing, Z.; Qi, R.; Thibonnier, M.; Ren, P. Molecular dynamics study of the hybridization between RNA and modified oligonucleotides. *J. Chem. Theory Comput.* **2019**, *15*, 6422−6432.
(27) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(28) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(29) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435−13441.
(30) Bonomi, M.; Parrinello, M. Enhanced sampling in the well-tempered ensemble. *Phys. Rev. Lett.* **2010**, *104*, 190601.
(31) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17874−17879.
(32) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 2179−2183.
(33) Dickerson, R. E.; Drew, H. R. Structure of a B-DNA dodecamer: II. Influence of base sequence on helix structure. *J. Mol. Biol.* **1981**, *149*, 761−786.
(34) Drew, H. R.; Dickerson, R. E. Structure of a B-DNA dodecamer: III. Geometry of hydration. *J. Mol. Biol.* **1981**, *151*, 535−556.
(35) Marky, L. A.; Blumenfeld, K. S.; Kozlowski, S.; Breslauer, K. J. Salt-dependent conformational transitions in the self-complementary deoxydodecanucleotide d (CGCAATTCGCG): Evidence for hairpin formation. *Biopolymers* **1983**, *22*, 1247−1257.
(36) SantaLucia, J., Jr; Hicks, D. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415−440.
(37) Owczarzy, R.; You, Y.; Moreira, B. G.; Manthey, J. A.; Huang, L.; Behlke, M. A.; Walder, J. A. Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry* **2004**, *43*, 3537−3554.
(38) Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(39) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D., Jr Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348−362.
(40) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D., Jr Impact of 2′-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929−1943.
(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(42) MacKerell, A. D., Jr; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(43) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM force field in GROMACS: analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.* **2010**, *6*, 459−466.
(44) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(45) Gil-Ley, A.; Bussi, G. Enhanced conformational sampling using replica exchange with collective-variable tempering. *J. Chem. Theory Comput.* **2015**, *11*, 1077−1085.
(46) Baftizadeh, F.; Cossio, P.; Pietrucci, F.; Laio, A. Protein folding and ligand-enzyme binding from bias-exchange metadynamics simulations. *Current Physical Chemistry* **2012**, *2*, 79−91.
(47) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255−268.
(48) Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695.

https://dx.doi.org/10.1021/acs.jpcb.0c09237
J. Phys. Chem. B 2021, 125, 771−779

(49) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(50) Nosé, S.; Klein, M. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055−1076.
(51) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(52) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(53) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(54) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R.; et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(55) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(56) Best, R. B.; Mittal, J. Balance between α and β structures in ab initio protein folding. *J. Phys. Chem. B* **2010**, *114*, 8790−8798.
(57) Guo, Z.; Thirumalai, D. Kinetics and thermodynamics of folding of ADE novodesigned four-helix bundle protein. *J. Mol. Biol.* **1996**, *263*, 323−343.
(58) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.
(59) Levy, Y.; Wolynes, P. G.; Onuchic, J. N. Protein topology determines binding mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 511−516.
(60) Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham, T., III; Darden, T.; Duke, R.; Gohlke, H.; et al. *Amber14*; University of California: San Francisco, CA, 2012.
(61) Langridge, R.; Marvin, D.; Seeds, W.; Wilson, H.; Hooper, C.; Wilkins, M.; Hamilton, L. The molecular configuration of deoxyribonucleic acid: II. Molecular models and their fourier transforms. *J. Mol. Biol.* **1960**, *2*, 38−IN12.
(62) Tiwary, P.; Parrinello, M. A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **2015**, *119*, 736−742.
(63) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. Extracting bulk properties of self-assembling systems from small simulations. *J. Phys.: Condens. Matter* **2010**, *22*, 104102.
(64) Doye, J. P.; Ouldridge, T. E.; Louis, A. A.; Romano, F.; Šulc, P.; Matek, C.; Snodin, B. E.; Rovigatti, L.; Schreck, J. S.; Harrison, R. M.; et al. Coarse-graining DNA for simulations of DNA nanotechnology. *Phys. Chem. Chem. Phys.* **2013**, *15*, 20395−20414.
(65) Ouldridge, T. E.; Šulc, P.; Romano, F.; Doye, J. P.; Louis, A. A. DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Res.* **2013**, *41*, 8886−8895.
(66) Mergny, J.-L.; Lacroix, L. Analysis of thermal melting curves. *Oligonucleotides* **2003**, *13*, 515−537.
(67) Yin, Y.; Zhao, X. S. Kinetics and dynamics of DNA hybridization. *Acc. Chem. Res.* **2011**, *44*, 1172−1181.
(68) Pfaendtner, J.; Bonomi, M. Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 5062−5067.
(69) Harvey, J. D.; Jena, P. V.; Baker, H. A.; Zerze, G. H.; Williams, R. M.; Galassi, T. V.; Roxbury, D.; Mittal, J.; Heller, D. A. A carbon nanotube reporter of microRNA hybridization events in vivo. *Nature biomedical engineering* **2017**, *1*, 0041.
(70) Bottaro, S.; Banáš, P.; Sponer, J.; Bussi, G. Free energy landscape of GAGA and UUCG RNA tetraloops. *J. Phys. Chem. Lett.* **2016**, *7*, 4032−4038.
(71) Saha, R.; Verbanic, S.; Chen, I. A. Lipid vesicles chaperone an encapsulated RNA aptamer. *Nat. Commun.* **2018**, *9*, 2313.

https://dx.doi.org/10.1021/acs.jpcb.0c09237
J. Phys. Chem. B 2021, 125, 771−779