THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2019, 123, 1939−1949

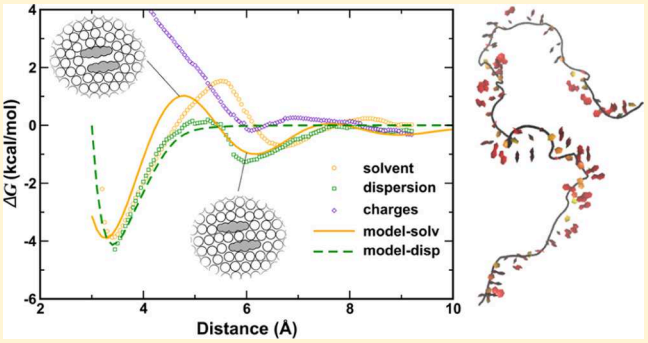
pubs.acs.org/JPCB

Theoretical Model for Solvent-Induced Base Stacking Interactions in Solvent-Free DNA Simulations

Chi H. Mak*

Departments of Chemistry and Biological Sciences, and Center of Applied Mathematical Sciences, University of Southern California, Los Angeles, California 90089, United States

ABSTRACT: Ultrahigh-throughput conformational sampling of biopolymers like nucleic acids are most effectively carried out without explicit solvents, but the physical origins of almost all inter- and intramolecular interactions controlling nucleic acid structures are rooted in water. Single-stranded (ss) DNAs or RNAs in water are characterized by ensembles of diverse conformations. To properly capture solvent-induced nucleobase stacking interactions in an otherwise solvent-free Monte Carlo algorithm, theoretical models are developed here to describe the solvent entropy and dispersion terms in base stacking free energies. To validate these models, equilibrium ensembles of ss $(dA)_n$ and $(dT)_n$ sequences (n = 30, 40, and 50) were simulated, and they quantitatively reproduced experimental small-angle X-ray scattering (SAXS) data. Simulated dA ensembles show substantial stacking. While less prevalent, stacking in dT chains is not negligible. Analysis of SAXS profiles suggests that excess features between wavevector 0.03 and 0.18 Å^{-1} correlate with stacking, and stacking in dA versus dT chains is chain length-dependent, where $(dT)_{30}$ and $(dA)_{30}$ chains have more similar structures, but longer dA chains show more stacking over dT. The average stack length in ss-dA chains is 5−10 nucleotides, yielding an estimate for the overall AlA stacking free energy at ∼1 kcal/mol.



1. INTRODUCTION

Base stacking is a key determinant of DNA stability, but the physical origin of stacking interactions is still debated. For a long time, $\pi-\pi$ interactions between delocalized electrons in the conjugated rings of purine and pyrimidine bases were thought to be the major component of stacking interactions. Evidence for this came mainly from ab initio electronic structure calculations of flat conjugated molecules, such as benzene, in vacuum.[1] However, more recent attempts to include solvent fields in ab initio calculations reveal that $\pi-\pi$ interactions produce very minimal effects that could not otherwise be attributed to dispersion, electrostatics, and solvent-induced forces,[2] and conventional molecular mechanics force fields like Amber[3] are good descriptors of these effects. The dispersion forces attributable to $\pi-\pi$ interactions between conjugated molecules also turn out to be largely offset by base−solvent dispersion forces, and understanding stacking therefore necessitates a detailed understanding of the role of the solvent. Recently, a large-scale molecular dynamics simulation was carried out aimed at quantifying the solvent's contribution to DNA base stacking free energies, and the results suggest that solvent entropy alone could account for majority of the stacking free energy between DNA bases in water.[4] The overall magnitude of this solvent-induced entropically derived stacking force seems to be mild, only of the order of a few kcal/mol per stack. Experimental evidence concurs that while stacking is the major driving force of helix stability, the double-helix is only marginally stable,[5−9] and base complementarity interactions surprisingly contribute an almost negligible amount to stability. Evidence from DNA melting experiments[10] also confirms that the unstacking of a base pair from helix amounts to only a few kcal/mol of free energy. However, resolving theoretical predictions with experimental measurements remains challenging because the measured stacking free energy is convoluted with other free energy terms such as the sugar−phosphate backbone conformational free energy, base complementarity interactions, and counterion-induced intrachain attractions, making the unambiguous separation of the free energy into its individual terms rather difficult.

In a recent paper, we have reported a new Monte Carlo (MC) simulation algorithm for single-stranded DNA and RNA.[11] To achieve high ergodicity, this algorithm makes use of an analytical formula to describe the conformational entropy of the sugar−phosphate backbone, summing over a large set of backbone conformational states using a mixed numerical-analytical scheme to eliminate the need to explicitly sample the backbone atoms. The analytical part of this "backbone conformational volume" (BCV) algorithm was formulated using inverse kinematics,[12,13] borrowing from well-established mathematics in robotics.[14] Refinement of this backbone closure formula by excluding conformations in which backbone

Received: November 7, 2018
Revised: February 2, 2019
Published: February 6, 2019

DOI: 10.1021/acs.jpcb.8b10848
J. Phys. Chem. B 2019, 123, 1939−1949

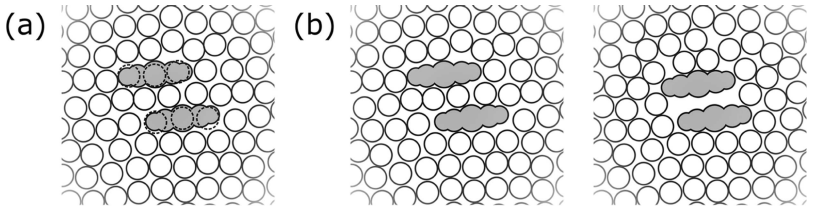


Figure 1. Sideview schematic showing two stacked bases (gray) solvated in water (white circles). (a) Solute molecules take up the space that was originally occupied by the water molecules indicated by the dashed lines. Expelling these solvent molecules causes an entropic penalty for the solvent. (b) As the separation between the two bases changes, the structural granularity of the solvent is felt by the solutes when the space between them is too small to accommodate the solvent, leading to an extra entropic overhead that forces the gap between them to close, driving the bases to stack.

atoms suffer any steric clash led to the final BCV/-c algorithm.[11,15] We have validated this BCV/-c algorithm by comparing MC simulation results to a number of experimental structural measurements on ssDNA and RNA.[11] Using the BCV/-c algorithm, large thermal ensembles of ssDNA and RNA chains can be simulated rapidly, and the technique has been applied to predicting single-stranded RNA secondary- and tertiary-structure thermodynamics, demonstrating strong agreements with experiments.[11,15]

Bases interact with each other in the original formulation of BCV/-c via steric interactions only. This was done deliberately so the BCV/-c algorithm could be validated against experiments without contamination from other base−base free-energy terms and a number of DNA and RNA constructs, primarily pyrimidine sequences (poly-dT or poly-U) where stacking is minimal, have been simulated and used to validate against experiments. To further extend the applicability of BCV/-c to sequences beyond these constructs, an accurate theoretical model for stacking free energies must be developed and integrated into the algorithm. This is the focus of this paper.

BCV/-c has been designed to be a solvent-free simulation. The key to high-ergodicity conformational sampling of solution-phase biopolymers is to not employ explicit water molecules, which significantly impede sampling efficiency. However, if base stacking forces are indeed driven by the solvent, finding a way to describe the solvent-induced effects without explicitly including the solvent requires a theory. The availability of solvent-free theoretical models for the different classes of molecular forces important for the accurate description of nucleic acid structures is therefore central to the success of the general strategy adopted by solvent-free simulations like BCV/-c MC. Previously, we have formulated a theoretical model for capturing the intra- and interchain free energies mediated by Mg^{2+} or Na^{+} counterions in the solution on the sugar−phosphate backbone into BCV/-c MC.[16−18] This paper is aimed at extending this development to provide a theoretical model for the solvent-induced base-stacking interactions.

2. PHYSICAL BASIS FOR THE SOLVENT'S ROLES IN BASE STACKING

Figure 1a shows a cartoon of two bases with their rings parallel to each other and solvated in water. The solvent molecules are colored white and the flat bases are colored gray. This drawing takes a sideview perspective—the two bases would be stacked when viewed from the top. In this drawing, there is a layer of solvent molecules lining the vertical separation between the two bases. At other separations, there would be different numbers of solvent molecules between them.

To accommodate a solute, the solvent sacrifices entropy. Figure 1a shows the solvent molecules that were originally in the space now occupied by the gray solute molecules using dashed lines. Solvating the solutes requires these solvent molecules be expelled from the space now occupied by solutes, and this incurs an entropic cost for the solvent. Recent studies of the solvation of proteins in water and the driving force of hydrophobic assemblies[19−27] suggest that the probability of observing a certain number of solvent molecules N in a cavity with volume V within the solvent is well described by a Gaussian theory when V is not too large

$$P_V(N) = (2\pi\chi_V)^{-1/2} \exp[-(N - \langle N\rangle_V)^2/2\chi_V] \quad (1)$$

where $\langle N\rangle_V = \rho V$ is the equilibrium value of N and ρ is the number density of the solvent, and

$$\chi_V = \langle(\delta N)^2\rangle_V = \rho V + \rho^2 \int_V d\mathbf{r} \int_V d\mathbf{r}' h(|\mathbf{r} - \mathbf{r}'|) \quad (2)$$

gives the second moment in the fluctuations of N, where $h(r) \equiv g(r) - 1$ and $g(r)$ are the radial distribution function of the neat liquid solvent. The free energy of observing no solvent molecule in the cavity then gives the approximate solvation free energy

$$\Delta G \approx k_B T(\rho V)^2/2\chi_V + k_B T \ln(2\pi\chi_V)/2 \quad (3)$$

where k_B is Boltzmann's constant and T is absolute temperature. Because $[g(r) - 1]$ cuts off quickly beyond the correlation length of the liquid, the free energy ΔG is roughly proportional to V. In a dense liquid where the repulsive part of the interactions drives structures, this ΔG is largely entropic in origin. As a rough estimate, expelling one water molecule at its normal density costs ∼4 kcal/mol of free energy at room temperature.

When two bases are far apart, the total solvation free energy of the two bases is expected to be the sum of each one individually. However, when the two bases approach each other, at some distances, the intrinsic structure of the solvent will begin to be felt by the bases. This is illustrated by Figure 1b. On the left, Figure 1b shows a scenario where the bases have a single layer of the solvent molecule between them. On the right, it shows another scenario where the two bases are slightly closer to each other, and the space between them is now too tight to be able to accommodate solvent molecules. At this separation, the two bases together exclude more water molecules that sum of them individually, and the solvent experiences an extra entropic penalty. The solvent's desire to eliminate this entropic overhead tends to force the bases

DOI: 10.1021/acs.jpcb.8b10848
J. Phys. Chem. B 2019, 123, 1939−1949

together to try to close this gap and constitutes the physical basis of the solvent-induced entropically-derived stacking force. When there are no solvent molecules between the two bases, their composite volume is somewhat smaller than the sum of them individually. This produces a stabilization force on two stacked bases.

These solvent entropic effects have been illustrated vividly in a recently reported large-scale explicit water simulation, where two bases (A|A or G|G, with no sugar−phosphate backbone) were stacked in a box of 10 000 TIP3P water, and the free energy of stacking was computed using umbrella sampling. The results are reproduced in Figure 2 for a G|G stack constrained

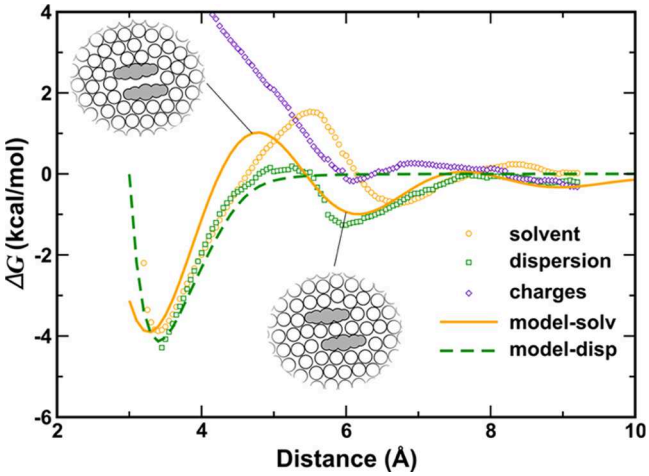


Figure 2. Data from the explicit-water simulation study in ref 4, showing the solvent entropic term (orange circles) in the stacking free energy between two guanines, as well as the dispersion term (green squares) and the electrostatic term (purple diamonds). Lines show theoretical models for the solvent entropic term (solid orange) and the dispersion term (dashed green) in the stacking free energy. The insets show schematics of how the solvent molecules are arranged around the bases at the first maximum and second minimum of the solvent entropic term (see discussion surrounding Figure 1).

to stay parallel to each other in the geometry that they would have if the stack was part of a B-DNA, with the x-axis being their vertical separate distance. The computed solvent-induced entropic part of the computed free energy is shown by the orange circles. This entropic part of the stacking free energy exhibits a striking oscillatory dependence on the base−base distance, which is most certainly a consequence of the solvent's granular structure. By applying perturbations to the intermolecular potentials, the simulation was also able to estimate the contributions of dispersion forces and partial-charge interactions between the two bases on the stacking free energy, and these are shown in Figure 2 as the green squares and the purple diamonds, respectively. The details for this perturbation procedure were described in ref 4, which involved calculating perturbations in the free energy when different types of energy terms in the intermolecular potentials were added successively to the system's total energy, starting from a system with no dispersion interactions and no electrostatics (i.e., with all van der Waals potentials replaced by the Weeks−Chandler−Andersen model[28] and no partial charges, giving the orange curve in Figure 2), to one adding hydrogen bonding in the solvent (the purple curve in Figure 2), and to one adding electrostatic interactions between the two bases (the green curve in Figure 2). Because these perturbations are not formally small in magnitude, the results are sensitive to the order in which these "perturbation" were applied, and any numerical uncertainties in the result from one perturbation will also propagate to the next. Nonetheless, the results in Figure 2 do show that the net perturbations to the free energy when successive potential energy terms are added to the system are quite small (no more than a few kcal/mol), even though the perturbation to the total potential energy itself may not be, and it lends support to the validity of this perturbative decomposition of the stacking free energy. The solvent-renormalized charge−charge interactions between the two Gs (purple diamonds) turn out to be effectively repulsive. On the other hand, the solvent-renormalized dispersion forces (green squares) are much weaker than the van der Waals interaction of two naked Gs without water. The range of both the dispersion and electrostatics terms of the stacking free energy are also shorter than without water, effectively cutting off at just around 6 Å. While both show a slight dip around 6 Å, this feature is likely due to an imprecision in how the free-energy terms were divided up based on the simulation data. Similarly, the solvent free-energy term shown by the orange circles are probably somewhat right-shifted also because of this. These results show that the solvent renormalizes the dispersion interaction between the two solutes heavily. Apparently, the loss of dispersion interaction between the two bases, as they separate from each other, is quickly offset by the gain in dispersion interactions between the bases and the solvent molecules as they refill the space between the bases, resulting in what appears to be a highly attenuated dispersion term in the stacking free energy.

3. MODELING THE SOLVENT ENTROPY TERM IN BASE STACKING FREE ENERGY

To develop a theory for the solvent entropic contribution to the stacking free energy, we start with eq 3. χ_V, the variance of the solvent density within the combined volume V occupied by the two bases, appears in both terms on the righthand side. Because the radial distribution function $g(r)$ for water is known, the calculation of χ_V is in principle straightforward for V of any shape and size. The most important part of the theory is how χ_V changes with the separation between the two bases. As a reference, one can take χ_V^0 to be the value of χ_V when the bases are far apart

$$\chi_V^0 = \rho V_A + \rho^2 \int_{V_A} d\mathbf{r} \int_{V_A} d\mathbf{r}' h(|\mathbf{r} - \mathbf{r}'|) + \rho V_B + \rho^2 \int_{V_B} d\mathbf{r} \int_{V_B} d\mathbf{r}' h(|\mathbf{r} - \mathbf{r}'|) \quad (4)$$

where V_A and V_B are the volumes occupied by the individual bases. At large separations, there are no cross terms in the double integrals because $h(r) \to 0$. However, as the two bases come closer to each other, cross terms begin to appear, leading to

$$\chi_V = \rho(V_A + V_B) + \rho^2 \int_{V_A+V_B} d\mathbf{r} \int_{V_A+V_B} d\mathbf{r}' h(|\mathbf{r} - \mathbf{r}'|) \quad (5)$$

The difference between eqs 4 and 5 gives $\Delta\chi_V = \chi_V - \chi_V^0$, where

$$\Delta\chi_V = 2\rho^2 \int_{V_A} d\mathbf{r} \int_{V_B} d\mathbf{r}' h(|\mathbf{r} - \mathbf{r}'|) \quad (6)$$

DOI: 10.1021/acs.jpcb.8b10848
J. Phys. Chem. B 2019, 123, 1939−1949

For the sizes and shapes of the purine and pyrimidine bases, the magnitude of $\Delta\chi_V$ is typically much smaller than χ_V^0. Therefore, when χ_V is substituted back into eq 3, ΔG should be well approximated by Taylor expanding both terms on the right, yielding, to linear order, the stacking free energy

$$\Delta G_{AB}(\mathbf{r};\varphi,\theta,\psi) - \Delta G_{AB}(\infty) \approx C\Delta\chi_V \tag{7}$$

where $\Delta G_{AB}(\mathbf{r};\varphi,\theta,\psi)$ is the solvent-entropy-induced stacking interaction between bases A and B at some vector distance $\mathbf{r}$ apart and with relative orientations specified by three Euler angles (φ,θ,ψ), and C is a constant coefficient independent of the identities of the bases. Notice that the Taylor expansion of the first term in eq 3 yields a first-order term with a negative coefficient, whereas the second has a positive coefficient, but the first term usually dominates so the sign of the coefficient C is overall negative. The variations in ΔG with distance and orientation are of the order of just a few kcal/mol per stack, rendering the Taylor expansion to linear order an accurate approximation. This expansion also removes the need of having to calculate χ_V^0, which is necessarily different for different collection of bases, and simplifies the theory significantly. Also, when there are more than two bases, eq 7 remains valid, except $\Delta\chi_V$ will have additional cross terms, and because of the structure of integrals involved, the additional cross terms are all pairwise. For example, if there are multiple bases A, B, ... Z, eq 6 becomes

$$\Delta\chi_V = 2\rho^2 \sum_{\alpha>\beta=A}^{Z} \int_{V_\alpha} d\mathbf{r} \int_{V_\beta} d\mathbf{r}' h(|\mathbf{r}-\mathbf{r}'|) \tag{8}$$

While eq 8 can be used directly in a simulation to estimate the stacking free energy, the calculation of the double integrals during runtime can be further simplified by breaking down the volume of each molecule α into a sum of nonintersecting spheres that tile that molecule's volume (or approximately as a sum of barely intersecting spheres). A natural choice is to place the centers of these spheres on the heavy (non-hydrogen) atoms in each base, and the volume of each base then becomes a sum over these spheres, turning eq 8 into

$$\Delta\chi_V = 2\rho^2 \sum_{\alpha>\beta=A}^{Z} \sum_{i\in\alpha} \sum_{j\in\beta} \int_{V_i} d\mathbf{r} \int_{V_j} d\mathbf{r}' h(|\mathbf{r}-\mathbf{r}'|) \tag{9}$$

where i and j are spheres on bases α and β, respectively, and the integrals in eq 8 then break up into a sum of smaller integrals, each involving a pair of spheres from two different bases. Because the bond distances between heavy atoms in the purine and pyrimidine bases vary from roughly 1.32 to 1.55 Å, we approximate each base by a sum of barely intersecting spheres with diameter D = 1.52 Å, centered at the positions of its heavy atoms. The double integral in eq 9 over the interiors of two spheres with distance r between their centers can be transformed to a two-dimensional integral

$$\begin{aligned} V_iV_jk(r) &= \int_{V_i} d\mathbf{r} \int_{V_j} d\mathbf{r}' h(|\mathbf{r}-\mathbf{r}'|) \\ &= (2\pi)^2 \int_{r-R}^{r+R} ds \left(1 - \frac{r^2+s^2-R^2}{2rs}\right) s^2 \\ &\quad \int_{s-R}^{s+R} dt \left(1 - \frac{s^2+t^2-R^2}{2st}\right) t^2 h(t) \end{aligned} \tag{10}$$

where $R = D/2$ is the radius of spheres V_i and V_j. Using $h(r)$ for water, the integration in eq 10 can be performed and stored numerically as a function of the distance r between the two spheres. Figure 3 shows $h(r)$ for water as the orange circles,

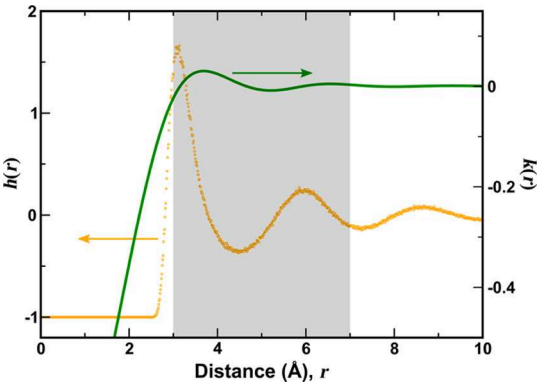


Figure 3. Radial distribution function $h(r) = g(r) - 1$ of water (orange circles),[4] and the function $k(r)$ defined in eq 10.

and the result for $k(r)$ is shown in as the solid green line. The range of distances where 3 Å < r < 7 Å is most relevant, because steric repulsion prevents two nonbonded non-hydrogen atoms from approaching distance less than ~3 Å, and $k(r)$ vanishes beyond ~7 Å. For example, if two spheres representing two nonbonded atoms are 3 Å apart, their ΔG would be repulsive because $k(r$ = 3 Å) is negative and the coefficient C in eq 3 is also negative. However, this ΔG would become attractive at $r \approx$ 3.2 Å because $k(r)$ changes sign there. This ΔG would then reach a minimum at $r \approx$ 3.7 Å, after which it switches sign again at $r \approx$ 4.3 Å. Using $k(r)$, eq 9 for any collection of bases simplifies to a sum over pairwise interactions between spheres centered on the heavy atoms belonging to different pairs of bases

$$\Delta\chi_V = 2\rho^2 \left(\frac{4\pi R^3}{3}\right)^2 \sum_{\alpha>\beta=A}^{Z} \sum_{i\in\alpha} \sum_{j\in\beta} k(r_{ij}) \tag{11}$$

where r_{ij} is the distance between sphere i on base α and sphere j on base β.

Applying this model to the system studied by the simulations summarized in Figure 2, the results of using eq 9 and $k(r)$ to compute the stacking free energy between a GlG stack are shown by the solid orange line in Figure 2. Compared to the simulation results in the open circles, the theoretical model predicts an analogous oscillatory behavior, manifestly the result of the solvent's inherent structural granularity, according to the peaks and valleys in $h(r)$. The magnitude of the coefficient C in eq 3 can be adjusted to reproduce the magnitude of the stacking free energy from the simulations. The oscillations in ΔG from theory are somewhat left-shifted compared to the simulations. Because the free-energy terms computed by the simulations were sensitive to how and the order in which the perturbations were applied, there is also an inherent uncertainty in how the free energy is decomposed into individual terms. Some of the discrepancy between the exact positions of the peaks and valleys in ΔG between the model and simulations can be partly attributed to this imprecision. Note that this uncertainty also introduces similar errors into the estimation of the dispersion term (green

DOI: 10.1021/acs.jpcb.8b10848
J. Phys. Chem. B 2019, 123, 1939−1949

squares) as well as the charge term (purple diamonds) because the terms are all correlated with each other.

4. MODELING SOLVENT-RENORMALIZED DISPERSION FORCE BETWEEN BASES

Next, we consider the dispersion term in the stacking free energy. Our model for the dispersion term is more empirical in nature compared to the entropy model. It starts with the supposition that besides just taking up the space that was originally occupied by the solvent molecules, the solutes do not significantly alter the structure of the solvent. This is clearly true in the limit where the solute molecules are small compared to the solvent molecules as well as in the opposite limit, where the solute molecules are large compared to the solvent molecules. In the regime where the solute and solvent molecules are comparable, the shape of the solute molecules does matter. The solute molecules could also polarize the solvent molecules to distort its structure away from the neat liquid, and these polarization effects are expected to be more important for larger solutes that have a significant interfacial area with the solvent.[27,29] But interestingly, the supposition that the solutes do not significantly perturb the solvent structure is corroborated by the entropy model described above. The entropy model also presupposes that the original solvent structure is largely unperturbed in the presence of the solutes; otherwise, the model, which is based on the radial distribution of the neat liquid solvent, would not have been correct. The fact that the entropy model largely reproduces the essential features of the simulation results lends support to the underlying assumption that the solvent's structure is not very different.

Assuming the dispersion forces are pairwise, which is the case when the Lennard-Jones potential is used to model dispersion, the total potential energy of the neat liquid solvent can be written as

$$U = \left\langle \sum_{i>j} u(r_{ij}) \right\rangle = \int_V \mathrm{d}\mathbf{r} \int_V \mathrm{d}\mathbf{r}' \rho^2 g(|\mathbf{r} - \mathbf{r}'|) u(|\mathbf{r} - \mathbf{r}'|) \tag{12}$$

where the sum goes over all pairs of solvent molecules, u is the pairwise potential between them, the angled brackets represent an ensemble average, g is the radial distribution function, and V is the entire volume of the solvent. Equation 12 also works if we are computing the interaction of one region in the solvent V_1 with another region V_2, simply by replacing the double volume integral over V by one over V_1 and the other over V_2. For example, if one wants to compute the original dispersion interactions between all the solvent molecules that were displaced by one of the solute molecules in Figure 1a (let this be A) with the solvent molecules that were displaced by the other solute (let this be B), one simply applies eq 12 and uses the volume of the two solutes V_A and V_B instead

$$U^{SS} = \int_{V_A} \mathrm{d}\mathbf{r} \int_{V_B} \mathrm{d}\mathbf{r}' \rho^2 g(|\mathbf{r} - \mathbf{r}'|) u(|\mathbf{r} - \mathbf{r}'|) \tag{13}$$

Analogously, if one wants to compute the original dispersion interactions between the solvent molecules displaced by both solutes in Figure 1a with the rest of the solvent, one simply applies eq 12 and uses the composite volume of the two solutes $V_A + V_B$ for one of the integrals and the complementary volume $V - (V_A + V_B)$ for the other

$$U^{Ss} = \int_{V_A+V_B} \mathrm{d}\mathbf{r} \int_{V-(V_A+V_B)} \mathrm{d}\mathbf{r}' \rho^2 g(|\mathbf{r} - \mathbf{r}'|) u(|\mathbf{r} - \mathbf{r}'|) \tag{14}$$

This double integral can be further decomposed into six terms, representing double integrals over V_A and V, V_A and V_A, V_A and V_B, V_B and V, V_B and V_A, and V_B and V_B.

When the solvent molecules are replaced by solutes, the magnitudes of the dispersion interactions U^{SS} between the two regions defined by the solutes, as well as the dispersion interactions U^{Ss} between the regions defined by the solutes and the rest of the solvent, will change, but we assume that the forms (i.e., how the integrals in eqs 13 and 14 depend on the separation and orientation of the two bases) do not. This then provides an approximate model for the functional form of the solvent-renormalized dispersion interactions between two bases. Note that in both eqs 13 and 14, the only terms that change with the separation and the relative orientations of the two bases are due to double integrals over V_A and V_B. All the other terms are invariant. Therefore, the functional form of the solvent-renormalized dispersion interactions between two solutes A and B is necessarily characterized by a pairwise potential, which has been modified by the radial distribution function $g(r)$ of the solvent

$$U^{SS}_{disp} = C_0 \int_{V_A} \mathrm{d}\mathbf{r} \int_{V_B} \mathrm{d}\mathbf{r}' g(|\mathbf{r} - \mathbf{r}'|) u(|\mathbf{r} - \mathbf{r}'|) \tag{15}$$

where C_0 is an empirical coefficient. Because this effective dispersion interaction is now controlled by both the range of g as well as u, g essentially screens u, giving the solvent-renormalized dispersion a much shorter distance range compared to the original u.

Using the functional form in eq 15 and $g(r)$ from Figure 3, we computed the solvent-renormalized dispersion interaction between a G|G stack. The results are shown in Figure 2 as the green dashed line. The model agrees with the simulations depicted by the green squares quite accurately up to about 5.6 Å, after which the simulation data show a dip between 5.6 and 7.6 Å. There is no physical reason to expect the dispersion interactions between the two bases should increase at this distance range, unless the bases polarize the solvent and alters it structure. As discussed above, this dip in the simulated dispersion could be a result of the imprecise nature of how the individual terms of the stacking free energy were extracted from the simulations. Interestingly, the dip in the simulated dispersion interaction coincides with a dip in the stacking free energy predicted by the solvent entropy model describe above, shown in Figure 2 as the orange line, suggesting that they may have similar origins.

5. THREE-BODY TERMS IN THE SOLVENT-ENTROPY-INDUCED STACKING FREE ENERGY

The solvent-entropy-driven stacking free energy model derived above contains only two-body interactions, manifestly in its final form in eq 11. The pairwise nature of this solvent entropy term comes from the Taylor expansion of the free energy to first order in $\Delta\chi_V$. However, going to higher orders will not fundamentally change this because the lack of three-body or higher correlations in the model is a consequence of the Gaussian model assumed in eq 1. According to Wick's theorem,[30] all higher moments of a Gaussian distribution can be re-expressed in terms of just the covariance. Therefore, any

DOI: 10.1021/acs.jpcb.8b10848
J. Phys. Chem. B 2019, 123, 1939−1949

theory based on eq 1 cannot contain independent terms higher than two-body. As the simulation data will demonstrate below, while the Gaussian model is quite accurate for two solute molecules, deficiencies due to the lack of three-body terms are quite noticeable when multiple bases are present. Predominately, the lack of three-body terms in the model reveals itself in the simulations as a bias toward excessively compact folded structures. The highly dense core in the interior of these compact conformations were overstabilized by excessive attractive two-body stacking forces, leading to too many base−base contacts. To repair these deficiencies, missing three-body terms in the inherently Gaussian theory in eq 1 must be added back to the model.

There is no existing theory of how or which kinds of three-body interactions are most appropriate for base stacking. To model them, we employ the Kirkwood superposition approximation,[31,32] assuming that the three-body distribution function g_{ABC} between three objects A, B, and C could be approximately constructed from their two-body correlations g_{AB}, g_{BC}, and g_{CA}, via the combination $g_{ABC} \approx g_{AB}g_{BC}g_{CA}$. This superposition satisfies all the permutation symmetry requirements for the three-body distribution and can be constructed easily once the two-body base stacking interactions have been computed. Adding this to the stacking free energy, eq 11 is modified to

$$\Delta\chi_V = 2\rho^2(4\pi R^3/3)^2 \sum_{\alpha>\beta=A}^{Z} \left[\sum_{i\in\alpha}\sum_{j\in\beta} k(r_{ij})\right] - C' \sum_{\alpha>\beta>\gamma=A}^{Z} \left[\sum_{i\in\alpha}\sum_{j\in\beta} k(r_{ij})\right]\cdot\left[\sum_{i\in\beta}\sum_{j\in\gamma} k(r_{ij})\right]\cdot\left[\sum_{i\in\gamma}\sum_{j\in\alpha} k(r_{ij})\right] \quad (16)$$

where C' is an empirical coefficient. The sign of the added three-body term counteracts two-body forces in the direction of reducing excessive stacking in densely packed regions, where three-body base contacts are prevalent.

6. RESULTS

The theoretical models for the solvent entropy and the solvent-renormalized dispersion terms in the stacking free energy were integrated into a solvent-free BCV/-c MC simulation. To validate the model, we used simulated ensembles of two sets of ssDNA constructs to assess the effects of base stacking. The first set consisted of poly-dT chains of lengths 30, 40, and 50-nts. Thought to have only minimal stacking, dT chains serve as a baseline model. The second set consisted of poly-dA chains of lengths 30, 40, and 50-nts. In contrast with dT chains, dA chains are thought to be more strongly stacked. Ensembles of poly-dT and poly-dA ssDNA strands in solution have been studied in great detail by Pollack et al.[33,34] and by Herschlag et al.[35] using small-angle X-ray scattering (SAXS). These DNA constructs and the particular chain lengths used in the simulations were selected to match those for which high-quality SAXS data are available. These simulations provide a rigorous validation of both the theoretical interaction models as well as the simulation method itself. For each of the simulation reported below, an equilibrium of 100 000 to 200 000 independent conformations were sampled using the BCV/-c algorithm,[12,13] with the two-body and three-body solvent entropic terms and the dispersion forces described above appended to the free-energy function.

For the validation of the stacking models and the simulations, poly-dA and poly-dT chains were selected because they should have maximal difference in stacking. Indeed, experimental SAXS data[35] suggested that the stiffness of ssDNA sequences with chain lengths between 8 and 100 bases could be largely controlled by stacking, with poly-dA having much higher stiffness compared to poly-dT chains. This conclusion came from comparing the scaling exponents of poly-dA and poly-dT chains to steric-only simulations.[35] However, more recent SAXS data[34] suggest that this difference might be much subtler than previously thought, showing that the conformations of $(dT)_{30}$ and $(dA)_{30}$ in solution are surprisingly similar around physiologically relevant ionic strengths. While the differences in conformational characteristics of poly-dA and poly-dT chains are undoubtedly related to differences in their stacking propensities, the relationship between their structures as revealed by scattering experiments and their stacking may not be as straightforward as previously thought. An accurate stacking model together with an efficient conformational sampling algorithm should be able to reproduce SAXS data, but at the same time also shed some light on this subtle relationship between ssDNA structures and base stacking.

Figure 4a shows snapshots from $(dA)_{30}$ simulations where the two-body solvent entropy model, the solvent-renormalized dispersion model, and the empirical three-body solvent entropy term have been integrated into the BCV/-c MC algorithm. The simulations also included an effective counterion model, which had been calibrated to an approximately 100 mM Na^+ concentration.[18] The snapshots in Figure 4a are random samples, illustrating some of the typical features observed in the ensemble and how the bases along this $(dA)_{30}$ sequence are stacked. Unlike in a B-DNA, these stacked structures appear to be much more heterogenous in ssDNAs. Also, while the prevalence of stacking in these poly-dA chains is clear, stacked bases rarely persist over very long stretches on the sequence, the majority of them having stack lengths roughly between 5 and 10-nts. Figure 5 shows atomistic details of the sugar−phosphate backbone and the bases from conformations in the simulated ensemble.

The simulated $(dA)_{30}$ conformational ensemble consisted of close to 200 000 independent configurations, some of which have been depicted in Figure 4a. An ensemble-averaged scattering profile was computed using these configurations with the CRYSOL program.[36] The results are shown in Figure 6a as the open squares, with the experimental SAXS profile[34] shown as the solid line. The experimental SAXS scattering were reported in arbitrary intensity units, so the magnitudes of the SAXS data (solid line) have been scaled to provide the best overlap with the simulation data (open squares), and the proper basis of comparison between the experiment and simulation is the functional dependence of the scattering intensity $q \times I(q)$ as a function of the wavevector q. The agreement between the experiment and simulation is clearly quite strong. The other sets of simulation data in Figure 6a (closed circles and diamonds), which will be described below, correspond to two other sets of model parameters, but their intensities have not been scaled relative to the open squares, showing absolute intensity differences among different models.

DOI: 10.1021/acs.jpcb.8b10848
J. Phys. Chem. B 2019, 123, 1939−1949

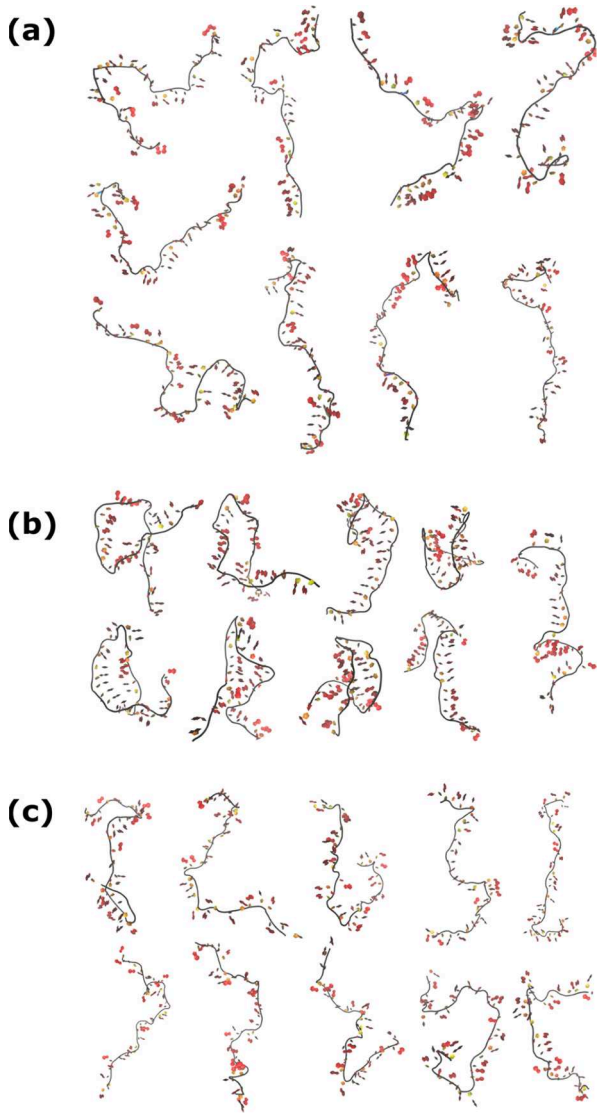


Figure 4. Snapshots of the $(dA)_{30}$ ensemble: (a) using all terms in the theoretical model (i.e., two- and three-body solvent entropic terms and solvent-renormalized dispersion), (b) omitting three-body solvent entropic term leads to overly compact structures, and (c) omitting solvent entropic terms all together leads to understacking of the bases.

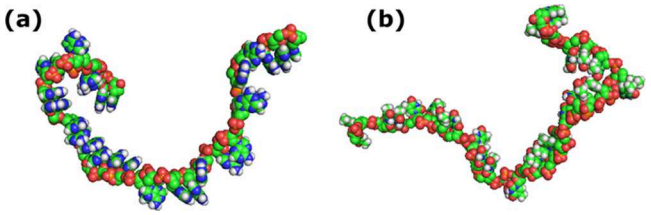


Figure 5. Space filling models of: (a) a $(dT)_{30}$ and (b) a $(dT)_{30}$ conformation from the simulated ensembles, showing greater atomistic details of the sugar−phosphate backbone and the bases.

To ascertain the effects of the two-body and three-body solvent entropy terms in the stacking free energy, Figure 4b shows snapshots from a $(dA)_{30}$ simulation without the three-body term, and Figure 4c shows snapshots from a simulation without any solvent entropy term at all. Comparing the conformations in Figure 4b to 4a, the chains appear to be much more compact without the three-body entropy term. The two-body-only solvent entropy model produces over-aggressive base−base interactions, and the incorporation of the three-body entropy term helps dissolve some of these excessively compact folds.

The quantitative effects of these terms in the stacking model on the scattering behavior of the chains are displayed in Figure 6a, which shows the ensemble-averaged SAXS profile corresponding to this simulation as the green diamonds. The lack of three-body interactions counteracting overaggressive two-body forces produced excessive structures in the wavevector range between 0.03 and 0.18 $Å^{-1}$, and these features correspond to some of the densely packed structures illustrated in Figure 4b. On the other hand, if there are no solvent-entropy-driven stacking forces at all (two- or three-body), stacking becomes quite minimal. In the absence of two- or three-body solvent entropy terms, the solvent-renormalized dispersion forces alone seem to be able to seed a small amount of stacking in $(dA)_{30}$, but its SAXS profile shown as the cyan circles in Figure 6a suggests that the chain is now understacked, revealed by a depletion of features in the wavevector range between 0.03 and 0.18 $Å^{-1}$.

In contrast to $(dA)_{30}$, $(dT)_{30}$ chains show a much lower level of stacking. This is illustrated in Figure 6b, which compares the experimental SAXS profile of $(dT)_{30}$ (solid line) to simulations with all terms in the stacking free turned on (open squares), simulations without the three-body solvent entropy term (green diamonds), and simulations without any solvent entropy term at all (cyan circles). In this case, the three simulated SAXS profiles are very similar, and they all seem to agree reasonably well with the experiment. These results corroborate the general assumption that dT chains have only minimal stacking. However, this is not to be taken as evidence that there is no stacking in dT chains at all. In fact, as we will show by snapshots from other simulations below, surprisingly, there appears to be a bit of stacking even in poly-dT chains. Also, the fact that the no-stacking simulation data depicted by the cyan circles in Figure 6b are lacking features near the peak of the profile compared to the experiment as well as to the other two sets of simulations also suggests that stacking in dT chains is not entirely negligible.

Going to 40-nt chains, Figure 7a shows simulation results for $(dT)_{40}$ and $(dA)_{40}$ scattering profiles compared to SAXS experimental data for $(dT)_{40}$.[33] Again, the simulations for $(dT)_{40}$ agree well with the experiment, although the experimental profile seems to contain features in the wavevector range from 0.10 to 0.18 $Å^{-1}$ not entirely captured by the simulations. Compared to $(dT)_{40}$, the simulations for $(dA)_{40}$ show an abundance of features in the wavevector range 0.03−0.18 $Å^{-1}$, suggesting stronger stacking in $(dA)_{40}$ compared to $(dT)_{40}$, as expected. Figure 7b shows simulations results for $(dT)_{50}$ and $(dA)_{50}$ compared to experimental data for $(dA)_{50}$.[35] Again, the simulations for $(dA)_{50}$ agree well with the experiment.

Snapshots from the simulated $(dA)_{50}$ ensemble are shown in Figure 8. Similar to the $(dA)_{30}$ conformations in Figure 4a, $(dA)_{50}$ shows clear signs of stacking. Compared to $(dA)_{30}$, the lengths of the stacks are not very different in $(dA)_{50}$. Again, the average stack length ∼5−10-nts. Figure 9 shows snapshots from the simulated $(dT)_{50}$ ensemble. A casual inspection of the $(dA)_{50}$ structures in Figure 8 and $(dT)_{50}$ in Figure 9 would not reveal any significant difference between the two ensembles. While stacking in $(dA)_{50}$ is noticeably more abundant, stacking

DOI: 10.1021/acs.jpcb.8b10848
J. Phys. Chem. B 2019, 123, 1939−1949

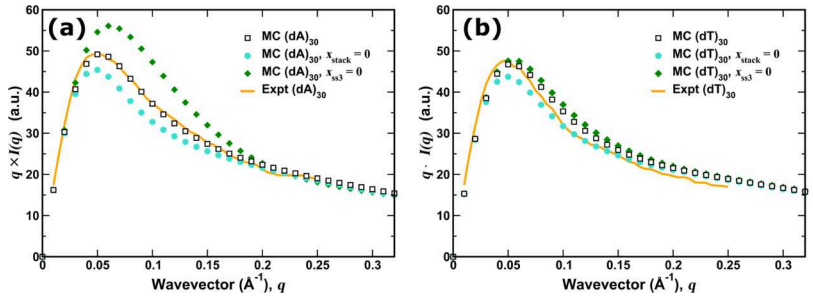


Figure 6. SAXS profiles for (a) $(dA)_{30}$ and (b) $(dT)_{30}$ chains, showing experimental data (solid line)[34] compared to MC simulation with the full stacking model (open squares), simulation without the three-body solvent entropy term (green diamonds) and simulations with no solvent entropy term all together (cyan circles).

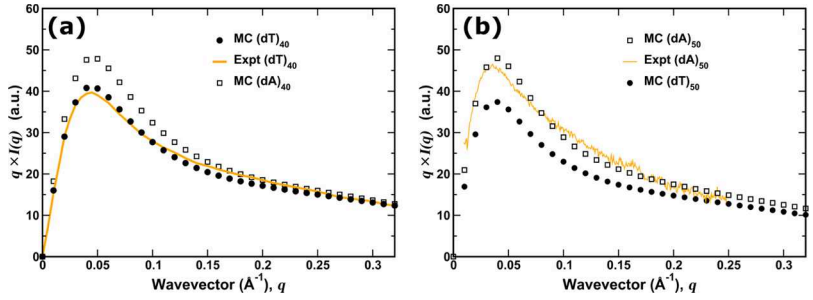


Figure 7. SAXS profiles for (a) 40-nt and (b) 50-nt poly-dT chains (solid circles) and poly-dA chains (open squares), showing experimental data (solid line)[33,35] compared to MC simulations.

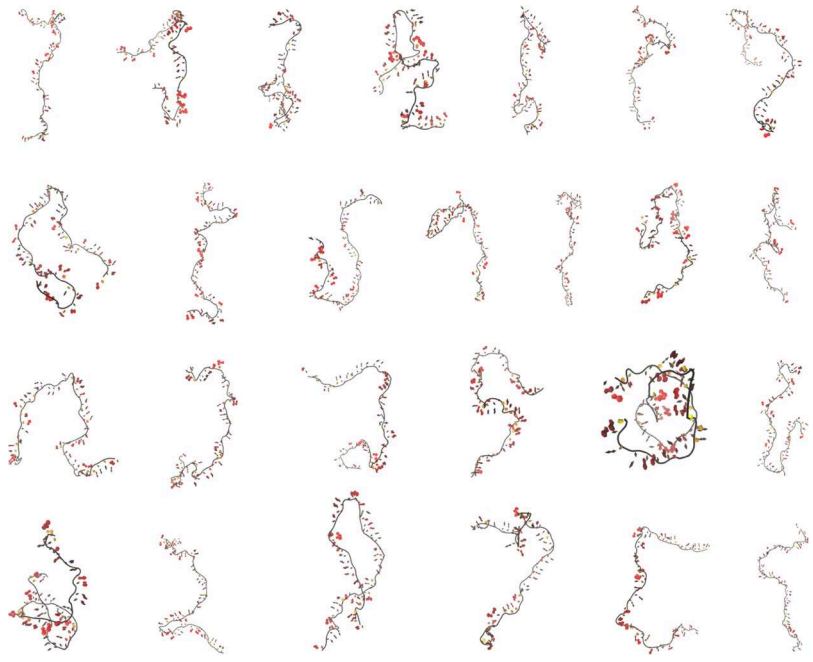


Figure 8. Snapshots from a simulated $(dA)_{50}$ ensemble.

in $(dT)_{50}$ is not negligible either. This apparent similarities in the conformations of poly-dA versus dT chains was also observed by Pollack et al.[34] based on their SAXS data. The SAXS profiles in Figure 7b, however, indicate that there are indeed differences in $(dA)_{50}$ compared to $(dT)_{50}$. A closer examination of the snapshots in Figures 8 and 9 indicates that $(dT)_{50}$ is somewhat more flexible, adopting more random coil structures more frequently. The persistence lengths along the backbone for both $(dA)_{50}$ and $(dT)_{50}$ chains turn out not to be very different, both ~3-nts; therefore, stacking does not seem to impact chain persistence and the inherent difference between $(dA)_{50}$ and $(dT)_{50}$ must be due to longer-range structures.

To understand the origin of the longer-lengthscale structures that differentiate $(dA)_n$ chains from $(dT)_n$ chains, consider a simple lattice model in which we map the nucleotide sequence onto one-dimensional space, and the stacking between two sequence-neighbor bases j and $j + 1$ is represented by a binary number s_j, where $s_j = 1$, if they are stacked and $s_j = 0$, otherwise. Let the free energy of two stacked bases relative to two

DOI: 10.1021/acs.jpcb.8b10848
J. Phys. Chem. B 2019, 123, 1939−1949

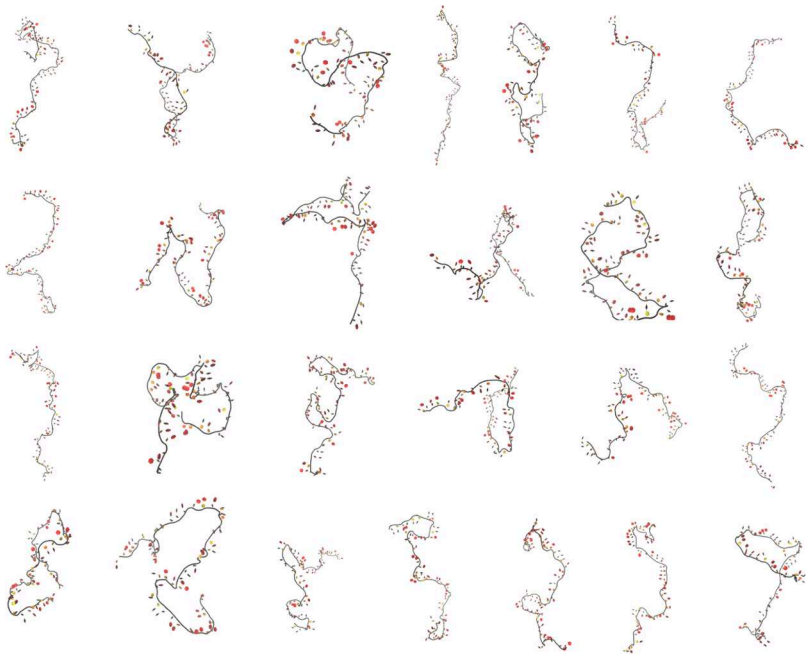


Figure 9. Snapshots from a simulated $(dT)_{50}$ ensemble.

unstacked bases be ϵ. For a sequence with n nts, the total free energy is then $F = \sum_{j=1}^{n-1} \epsilon s_j$. Summing over all stacking configurations along the sequence, the partition function of this system becomes $Z = (e^{-\beta\epsilon} + 1)^{n-1}$, where $\beta = 1/k_B T$, and the equilibrium average number of stacked base along the chain is simply $\langle \sum_j s_j \rangle = (n-1)p_1$, where $p_1 = e^{-\beta\epsilon}/(1 + e^{-\beta\epsilon})$ is the probability that a pair of sequence-neighbor bases are stacked. Clearly, there is no true long-range stacking order in this model as in the simulation results, but the mean stack length is well-defined. Using this model, one can easily compute the average length of a stacked sequence on this chain to be $\langle l \rangle = 1 + \frac{p_1}{1-p_1} \cdot [1 + p_1^{\,n}(np_1 - n - 1)] \xrightarrow[n\to\infty]{} (1-p_1)^{-1}$. (The same result can be derived assuming stacking follows Poisson statistics.) The average stack length in this model does not depend on the chain length as long as $n \gtrsim 30$. Using $\epsilon = -1$ kcal/mol, one obtains $\langle l \rangle \approx 6$ at 310 K, and if $\epsilon = -2$ kcal/mol, $\langle l \rangle \approx 27$. On the basis of this, our observation from the simulated ensembles that the average stack length in $(dA)_n$ chains is between 5 and 10-nts places the stacking energy between −0.85 and −1.4 kcal/mol. (This estimate for the stacking free energy not only includes effects from the direct stacking free energy between two bases described in the models above, but it also includes effects from the entropic costs of the sugar−phosphate backbone in the BCV/-v model to attain a conformation consistent with what is required for two sequence-neighbor bases to properly stack against each other.) This particular estimate that stacking is stabilized by 0.85−1.4 kcal/mol of free energy per stack is very much in line with experimental estimates, which show that for an A|A stack, the effective stacking free energy is ∼−1 kcal/mol at room temperature.[9]

The simple lattice model described here assumes the only relevant energy comes from the stacking of sequence-neighbor bases. Of course, stacking interactions may extend over longer separation in sequence space than this. For example, if the stacking of bases $j-1$ and j induces the stacking of bases j and $j+1$, next-nearest-neighbor interactions would enter the model. More likely though, because the chain can bend, the interaction between two bases far apart in sequence can also induce stacking in the local sequence neighborhood of each of the bases. In this case, it is possible that stacking could be seeded by long-range interactions along the sequence, and a careful measurement of the average stack length $\langle l \rangle$ as a function of the sequence length may provide further information on these long-range interactions.

Figure 10 shows a plot of the SAXS scattering intensities of poly-dA divided by poly-dT from the 30-, 40- and 50-nt

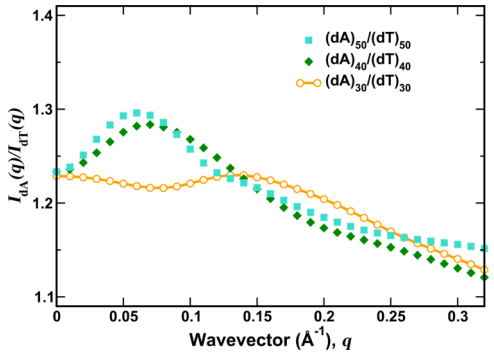


Figure 10. SAXS profile intensities for poly-dA chains divided by same-length poly-dT chains for 30-, 40- and 50-nt sequences from simulations. Excess features between $q = 0.03$ and 0.18 Å^{-1}, comparing dA against dT chains, reflect enhanced stacking.

ensembles in MC simulations. As described above, the features in the scattering profile between wavevector 0.03 and 0.18 Å^{-1} are related to enhanced stacking. Figure 10 shows that for $(dA)_{40}$ and $(dA)_{50}$, these features are more intense than in $(dT)_{40}$ (green diamonds) or $(dT)_{50}$ (cyan squares), suggesting that for these chain lengths, stacking is more abundant in poly-dA than poly-dT, as expected. However, in the 30-nt dA chains, these features are attenuated compared to $(dT)_{30}$ (orange circles in Figure 10) and the peak between 0.03 and

DOI: 10.1021/acs.jpcb.8b10848
J. Phys. Chem. B 2019, 123, 1939−1949

0.18 Å^{-1} is now missing, suggesting that the propensity of stacking in $(dA)_{30}$ and $(dT)_{30}$ are less dissimilar than in 40- or 50-nt chains. This supports the hypothesis that longer chains, because they are able to make more contacts between bases far apart in sequence, may be able to seed more stacking interactions in poly-dA chains over poly-dT, and stacking propensity acquires an apparent chain-length dependence.

7. CONCLUSIONS

An implicit-solvent analytical theory has been developed to model base stacking interactions in DNA. On the basis of results from previous simulation studies, indicating that the majority of the stacking free energy between DNA bases could be explained by solvent entropic effects, an analytic expansion of the entropic costs for forming a cavity in the solvent occupied by stacked bases have been derived as a function of positions and orientations of the bases involved. Supplemented with a theory for the solvent-renormalized dispersion energy as well as an empirical function for three-body stacking forces, this stacking model has been incorporated into a ssDNA simulation based on the BCV/-c algorithm previously reported, in order to construct a high-throughput simulation method for DNAs in solution without explicit solvent but correctly capturing the solvent's role in enforcing stacking interactions among the nucleobases. The results of the model and the simulations were validated against experimental data from SAXS of poly-dA and poly-dT chains with sequence lengths between 30 and 50 nucleotides. The simulation results quantitatively reproduced experimental SAXS data and they were able to provide insights on the nature of the subtle differences observed in the scattering profiles between poly-dA and poly-dT chains and how these differences are related to the different stacking propensities of A versus T. Simulated dA ensembles show substantial stacking. While less prevalent, stacking in dT chains is not negligible. Analysis of SAXS profiles suggests that excess features between wavevector 0.03 and 0.18 Å^{-1} correlate with stacking, and stacking in dA versus dT chains is chain length-dependent, where $(dT)_{30}$ and $(dA)_{30}$ chains have more similar structures, but longer dA chains show more stacking over dT. Average stack length in ss-dA chains is 5−10 nucleotides, yielding an estimate for the overall AlA stacking free energy at ∼1 kcal/mol. These results demonstrate that the BCV/-c algorithm is a viable numerical method for the conformational sampling of DNAs in solution, able to capture full solvent effects in a solvent-free simulation, making possible high-throughput simulations of nucleic acids that are "structureless", where their conformations must be characterized by a statistical thermal ensemble instead of a single minimum free-energy fold.

AUTHOR INFORMATION

Corresponding Author

*E-mail: cmak@usc.edu. Phone: 213-740-4101. Fax: 213-740-3972.

ORCID

Chi H. Mak: 0000-0002-5516-3304

Notes

The author declares no competing financial interest.

ACKNOWLEDGMENTS

This material is based upon the work supported by the National Science Foundation under grants nos. CHE-0713981 and CHE-1664801. We thank Adelene Sim and Dan Herschlag for providing us with SAXS raw data from ref 35.

REFERENCES

(1) Wheeler, S. E. Understanding Substituent Effects in Noncovalent Interactions Involving Aromatic Rings. *Acc. Chem. Res.* **2012**, *46*, 1029−1038.

(2) Šponer, J.; Šponer, J. E.; Mládek, A.; Jurečka, P.; Banáš, P.; Otyepka, M. Nature and Magnitude of Aromatic Base Stacking in DNA and Rna: Quantum Chemistry, Molecular Mechanics, and Experiment. *Biopolymers* **2013**, *99*, 978−988.

(3) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; 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.

(4) Mak, C. H. Unraveling Base Stacking Driving Forces in DNA. *J. Phys. Chem. B* **2016**, *120*, 6010−6020.

(5) Bommarito, S.; Peyret, N.; SantaLucia, J. Thermodynamic Parameters for DNA Sequences with Dangling Ends. *Nucleic Acids Res.* **2000**, *28*, 1929−1934.

(6) Kool, E. T. Hydrogen Bonding, Base Stacking, and Steric Effects in Dna Replication. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 1−22.

(7) Protozanova, E.; Yakovchuk, P.; Frank-Kamenetskii, M. D. Stacked-Unstacked Equilibrium at the Nick Site of DNA. *J. Mol. Biol.* **2004**, *342*, 775−785.

(8) SantaLucia, J., Jr.; Hicks, D. The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415−440.

(9) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. Base-Stacking and Base-Pairing Contributions into Thermal Stability of the DNA Double Helix. *Nucleic Acids Res.* **2006**, *34*, 564−574.

(10) 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.

(11) Mak, C. H. Atomistic Free Energy Model for Nucleic Acids: Simulations of Single-Stranded DNA and the Entropy Landscape of Rna Stem-Loop Structures. *J. Phys. Chem. B* **2015**, *119*, 14840−14856.

(12) Mak, C. H.; Matossian, T.; Chung, W.-Y. Conformational Entropy of the Rna Phosphate Backbone and Its Contribution to the Folding Free Energy. *Biophys. J.* **2014**, *106*, 1497−1507.

(13) Mak, C. H.; Sani, L. L.; Villa, A. N. Residual Conformational Entropies on the Sugar-Phosphate Backbone of Nucleic Acids: An Analysis of the Nucleosome Core DNA and the Ribosome. *J. Phys. Chem. B* **2015**, *119*, 10434−10447.

(14) Coutsias, E. A.; Seok, C.; Jacobson, M. P.; Dill, K. A. A Kinematic View of Loop Closure. *J. Comput. Chem.* **2004**, *25*, 510−528.

(15) Mak, C. H.; Phan, E. N. H. Topological Constraints and Their Conformational Entropic Penalties on Rna Folds. *Biophys. J.* **2018**, *114*, 2059−2071.

(16) Mak, C. H.; Henke, P. S. Ions and Rnas: Free Energies of Counterion-Mediated Rna Fold Stabilities. *J. Chem. Theory Comput.* **2012**, *9*, 621−639.

(17) Henke, P. S.; Mak, C. H. Free Energy of Rna-Counterion Interactions in a Tight-Binding Model Computed by a Discrete Space Mapping. *J. Chem. Phys.* **2014**, *141*, 064116.

(18) Henke, P. S.; Mak, C. H. An Implicit Divalent Counterion Force Field for Rna Molecular Dynamics. *J. Chem. Phys.* **2016**, *144*, 105104.

(19) Pratt, L. R.; Chandler, D. Theory of Hydrophobic Effect. *J. Chem. Phys.* **1977**, *67*, 3683−3704.

(20) Pratt, L. R.; Chandler, D. Hydrophobic Solvation of Nonspherical Solutes. *J. Chem. Phys.* **1980**, *73*, 3430−3433.

(21) Pratt, L. R.; Chandler, D. Hydrophobic Interactions and Osmotic 2nd Virial-Coefficients for Methanol in Water. *J. Solution Chem.* **1980**, *9*, 1−17.

DOI: 10.1021/acs.jpcb.8b10848
J. Phys. Chem. B 2019, 123, 1939−1949

(22) Wallqvist, A.; Berne, B. J. Computer-Simulation of Hydrophobic Hydration Forces on Stacked Plates at Short-Range. *J. Phys. Chem.* **1995**, *99*, 2893−2899.
(23) Garde, S.; Hummer, G.; García, A. E.; Paulaitis, M. E.; Pratt, L. R. Origin of Entropy Convergence in Hydrophobic Hydration and Protein Folding. *Phys. Rev. Lett.* **1996**, *77*, 4966−4968.
(24) Hummer, G.; Garde, S.; Garcia, A. E.; Pohorille, A.; Pratt, L. R. An Information Theory Model of Hydrophobic Interactions. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8951−8955.
(25) Lum, K.; Chandler, D.; Weeks, J. D. Hydrophobicity at Small and Large Length Scales. *J. Phys. Chem. B* **1999**, *103*, 4570−4577.
(26) Huang, D. M.; Geissler, P. L.; Chandler, D. Scaling of Hydrophobic Solvation Free Energies. *J. Phys. Chem. B* **2001**, *105*, 6704−6709.
(27) Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature* **2005**, *437*, 640−647.
(28) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role of Repulsive Forces in Determining Equilibrium Structure of Simple Liquids. *J. Chem. Phys.* **1971**, *54*, 5237−5247.
(29) Chandler, D. Hydrophobicity: Two Faces of Water. *Nature* **2002**, *417*, 491.
(30) Fetter, A. L.; Walecka, J. D. *Quantum Theory of Many-Particle Systems*; Dover Publications: Mineola, NY, 2003.
(31) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(32) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids: With Applications of Soft Matter*; Elsevier/AP: Amstersdam, 2013.
(33) Chen, H.; Meisburger, S. P.; Pabit, S. A.; Sutton, J. L.; Webb, W. W.; Pollack, L. Ionic Strength-Dependent Persistence Lengths of Single-Stranded Rna and DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *109*, 799−804.
(34) Plumridge, A.; Meisburger, S. P.; Andresen, K.; Pollack, L. The Impact of Base Stacking on the Conformations and Electrostatics of Single-Stranded DNA. *Nucleic Acids Res.* **2017**, *45*, 3932−3943.
(35) Sim, A. Y. L.; Lipfert, J.; Herschlag, D.; Doniach, S. Salt Dependence of the Radius of Gyration and Flexibility of Single-Stranded DNA in Solution Probed by Small-Angle X-Ray Scattering. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2012**, *86*, 021901.
(36) Svergun, D.; Barberato, C.; Koch, M. H. J. Crysol - a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28*, 768−773.

DOI: 10.1021/acs.jpcb.8b10848
J. Phys. Chem. B 2019, 123, 1939−1949