The self-assembly of DNA Holliday junctions studied with a minimal model

Thomas E. Ouldridge, Iain G. Johnston, Ard A. Louis, and Jonathan P. K. Doye

Citation: The Journal of Chemical Physics **130**, 065101 (2009);
View online: https://doi.org/10.1063/1.3055595
View Table of Contents: http://aip.scitation.org/toc/jcp/130/6
Published by the American Institute of Physics

Articles you may be interested in

Introducing improved structural properties and salt dependence into a coarse-grained model of DNA
The Journal of Chemical Physics **142**, 234901 (2015); 10.1063/1.4921957

Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model
The Journal of Chemical Physics **134**, 085101 (2011); 10.1063/1.3552946

A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality
The Journal of Chemical Physics **132**, 035105 (2010); 10.1063/1.3269994

Multifunctional energy landscape for a DNA G-quadruplex: An evolved molecular switch
The Journal of Chemical Physics **147**, 152715 (2017); 10.1063/1.4997377

CAMELOT: A machine learning approach for coarse-grained simulations of aggregation of block-copolymeric protein sequences
The Journal of Chemical Physics **143**, 243123 (2015); 10.1063/1.4935066

A nucleotide-level coarse-grained model of RNA
The Journal of Chemical Physics **140**, 235102 (2014); 10.1063/1.4881424

The self-assembly of DNA Holliday junctions studied with a minimal model

Thomas E. Ouldridge,[1] Iain G. Johnston,[1] Ard A. Louis,[1,a)] and Jonathan P. K. Doye[2,b)]
[1]*Rudolf Peierls Centre for Theoretical Physics, 1 Keble Road, Oxford OX1 3NP, United Kingdom*
[2]*Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QZ, United Kingdom*

(Received 21 July 2008; accepted 9 November 2008; published online 11 February 2009)

In this paper, we explore the feasibility of using coarse-grained models to simulate the self-assembly of DNA nanostructures. We introduce a simple model of DNA where each nucleotide is represented by two interaction sites corresponding to the sugar-phosphate backbone and the base. Using this model, we are able to simulate the self-assembly of both DNA duplexes and Holliday junctions from single-stranded DNA. We find that assembly is most successful in the temperature window below the melting temperatures of the target structure and above the melting temperature of misbonded aggregates. Furthermore, in the case of the Holliday junction, we show how a hierarchical assembly mechanism reduces the possibility of becoming trapped in misbonded configurations. The model is also able to reproduce the relative melting temperatures of different structures accurately and allows strand displacement to occur. © *2009 American Institute of Physics*. [DOI: 10.1063/1.3055595]

I. INTRODUCTION

The ability to design nanostructures which accurately self-assemble from simple units is central to the goal of engineering objects and machines on the nanoscale. Without self-assembly, structures must be laboriously constructed in a step by step fashion. Double-stranded DNA (dsDNA) has the ideal properties for a nanoscale building block,[1,2] with structural length scales determined by the separation of base pairs, the helical pitch, and its persistence length (approximately 0.33, 3.4,[3] and 50 nm,[4] respectively). Over these distances, dsDNA acts as an almost rigid rod and so it is capable of forming well-defined three-dimensional structures.

It is the selectivity of base pairing between single strands, however, that makes DNA ideal for controlled self-assembly. By designing sections of different strands to be complementary, a certain configuration of a system of oligonucleotides can be specified as the global minimum of the energy landscape. In this way, the target structure (usually consisting of branched double helices) can be "programed" into the sequences. This approach was initially demonstrated for a four-armed junction by the Seeman group in 1983.[5] Such junctions and more rigid double crossover motifs[6] can then be used to create two-dimensional lattices.[7,8] Yan *et al.*[9] also constructed ribbons and two-dimensional lattices from larger four-armed structures, each arm consisting of a junction of four strands. Furthermore, using Rothemund's DNA "origami" approach an almost arbitrary variety of two-dimensional shapes can be created.[10]

Progress in forming three-dimensional DNA nanostructures was initially much slower. The Seeman group managed to synthesize a DNA cube[11] and a truncated octahedron,[12] but only after a long series of steps and with a low final yield. More recently, approaches have been developed that allow polyhedral cages, such as tetrahedra,[13] trigonal bipyramids,[14] octahedra,[15,16] dodecahedra, and truncated icosahedra,[17] to be obtained in high yields simply by cooling solutions of appropriately designed oligonucleotides from high temperature. Additional structures have also been produced using preassembled modular building blocks incorporating other organic molecules.[18,19]

In designing strand sequences, it is important to minimize the stability of competing structures with respect to the stability of the target configuration. In addition, if systems can be designed to follow certain routes through configuration space—for example, by the hierarchical assembly of simple motifs[20]—the target can potentially be reached more efficiently. A standard approach to hierarchical assembly, such as that described by He *et al.*,[17] involves choosing sequences so that bonds between different pairs of oligonucleotides become stable at different temperatures. This allows certain motifs to form in isolation at high temperatures before bonding to each other as the solution is cooled. An alternative, elegant system for programming assembly pathways has been proposed by Yin *et al.*,[21] which relies on the metastability of single-stranded loop structures and the possibility of catalyzing their interactions using other oligonucleotides.

Given these recent experimental advances in creating DNA nanostructures, it would be useful to have theoretical models that allow further insights into the self-assembly process. In particular, a successful model would be able to provide information on the formation pathways and free energy landscape associated with the self-assembly, and as such would be of use to experimentalists wishing to consider increasingly more complex designs. Atomistic simulations of DNA would offer potentially the most spatially detailed descriptions of the self-assembly. However, they are computationally very expensive and are generally restricted to time scales that are too short to study self-assembly.[22]

Statistical approaches such as that of Poland and

a)Electronic mail: ard.louis@physics.ox.ac.uk.
b)Electronic mail: jonathan.doye@chem.ox.ac.uk.

 © 2009 American Institute of Physics

Scheraga[23] and the nearest-neighbor model[24] use simple expressions for the free energy of helix and random coil states to obtain equilibrium results for the bonding of two strands. While the parameters in these models can be tuned to give very accurate correspondence with experimental data,[25,26] they give no information on the dynamics and formation pathways and hence are only useful for ensuring that the target structure has significantly lower free energy than competing configurations. Furthermore, any description purely based on secondary structure (i.e., which bases are paired) is inherently incapable of accounting for topological effects such as linking of looped structures.[27]

Coarse-grained or minimal models offer a compromise between detail and computational simplicity, and are well suited to the study of hybridization of oligonucleotides. The aim of these models is to be capable of describing both the thermodynamic and kinetic behaviors of systems, a vital feature if kinetic metastability is inherent in assembly pathways.[21] In developing such minimal models, the approach is usually to retain just those physical features of the system that are essential to the behavior that is of interest.

The models of Dauxois *et al.*[28] and modified versions such as that proposed by Buyukdagli *et al.*[29] constitute the simplest class of dynamical models. Although these are dynamic models in the sense that the energy is a function of the separation between each base, the nucleotides are constrained to move in one dimension. This lack of conformational freedom means that these models are incapable of capturing the nuances of the self-assembly from single-stranded DNA (ssDNA).

Recently, models have been proposed which capture the helicity of dsDNA using two[30] or three[31] interaction sites per nucleotide. These models, however, are optimized for studying deviations from the ideal double-stranded state, and so have not been used to examine self-assembly. Although they have been used to study the thermal denaturation of dsDNA, it is essential for our purposes to be able to simulate the assembly of a structure from ssDNA as it is this process that will reveal the kinetic traps and free energy landscape associated with the formation of a particular DNA nanostructure.

Simpler, linear models, also with two interaction sites per nucleotide, have been used to investigate duplex hybridization,[32] hairpin formation,[33] and gelation of colloids functionalized with oligonucleotides.[34] These models all use two interaction sites to represent one nucleotide, with backbone sites linked to each other to represent the sugar-phosphate chain, and interaction sites which represent the bases. This work investigates the possibility of extending the use of such coarse-grained models to study the self-assembly of nanostructures that involve multiple strands forming branched duplexes. We hypothesize that the self-assembly properties of DNA are dominated by the fact that ssDNA is a semiflexible polymer with selective attractive interactions. We introduce an extremely simple model, similar to that of Starr and Sciortino, to test this hypothesis.[34] This simplicity enables us to explore the thermodynamics and kinetics of self-assembly in the model in great depth, and hence examine the feasibility of simulating nanostructure formation through minimal models.

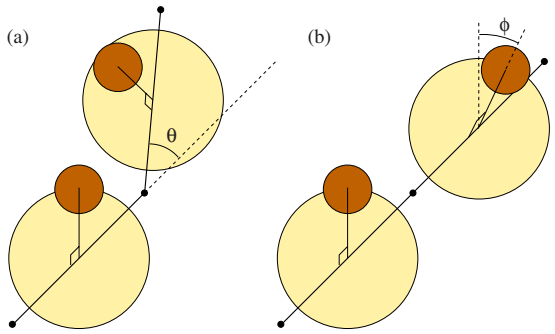

FIG. 1. (Color online) A schematic representation of the model. The thick lines represent the rigid backbone monomer units and the large circles represent the repulsive LJ interactions at their centers. The smaller, darker circles represent the bases. The panels illustrate the definitions of (a) the bending angle between two units (θ) and (b) the torsional angle (ϕ) which is found after the monomers have been rotated to lie parallel.

We first describe the model in Sec. II, and then examine its success in reproducing the general features of hybridization in Sec. III A. Next in Sec. III B, we apply it to the formation of a Holliday junction, a simple nanostructure consisting of a four-armed cross.[8]

II. METHODS

A. Model

We introduce an off-lattice model inspired by that which Starr and Sciortino used to study the gelation of four-armed DNA dendrimers.[34] As our aim is to reproduce the basic physics with as simple a model as possible, we neglect contributions to the interactions due to base stacking, and the charge and asymmetry of the phosphate backbone. We do not attempt to include the detailed geometrical structure of DNA but instead represent the oligonucleotides as a chain of monomer units, each corresponding to one nucleotide (Fig. 1). A monomer consists of a rod (chosen to be rigid for simplicity) of length l with a repulsive backbone interaction site at the center of the rod. In addition, each unit has a bonding interaction site (or base) at a distance of $0.3l$ from the backbone site (perpendicular to the rod). Each monomer is also assigned a base type (A, G, C, and T) to model the selective nature of bonding. In this model, we only consider bonds between the complementary pairs A-T and G-C.

We do not explicitly include any solvent molecules in our simulations but instead use effective potentials to describe the interactions between the DNA. Sites interact through shifted-force Lennard-Jones (LJ) potentials, where, as well as truncating and shifting the potential, an extra term is included to ensure the force goes smoothly to zero at the cutoff r_c. For $r < r_c$,

$$V_{\mathrm{sf}}(r) = V_{\mathrm{LJ}}(r) - V_{\mathrm{LJ}}(r_c) - (r - r_c)\left.\frac{dV_{\mathrm{LJ}}}{dr}\right|_{r=r_c}, \qquad (1)$$

where

$$V_{\mathrm{LJ}}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{2}$$

and $V_{\mathrm{sf}}(r)=0$ for $r \geq r_c$. Backbone sites (except adjacent units on the same strand) interact through Eq. (1) with $\sigma=l$ and $r_c=2^{1/6}\sigma$. This purely repulsive interaction models the steric repulsion between strands. Bonding sites (again excluding adjacent units on a strand) interact via Eq. (1) with $\sigma=0.35l$ and $r_c=2.5\sigma$ for complementary bases (to allow for attraction) and $r_c=2^{1/6}\sigma$ for all other pairings. The depth of the resulting potential well between complementary bases, $\epsilon_{\mathrm{base}}^{\mathrm{eff}}$, is 0.396ϵ. In what follows, we will measure the temperature in terms of a reduced temperature, $T^*=k_BT/\epsilon_{\mathrm{base}}^{\mathrm{eff}}$. The above choice of parameters ensures that the attractive interaction between complementary bases is largely shielded by backbone repulsion. Monomers therefore bond selectively and can only bond strongly to one other monomer at a given instant. These are the key features of Watson–Crick base pairing that make DNA so useful for self-assembly.

The model also includes potentials between consecutive monomers associated with bending and twisting the strand

$$V_{\mathrm{bend}} = \begin{cases} k_1(1-\cos(\theta)) & \text{if } \theta < \dfrac{3\pi}{4} \\ \infty & \text{otherwise,} \end{cases} \tag{3}$$

and

$$V_{\mathrm{twist}} = k_2(1-\cos(\phi)). \tag{4}$$

We define θ as the angle between the vectors along the adjacent monomer rods. As previously mentioned, consecutive backbone sites do not interact via LJ potentials. Instead, a hard cutoff is introduced in Eq. (3) to reflect the fact that an oligonucleotide cannot double back on itself. ϕ is taken as the angle between adjacent backbones to bonding site vectors after the monomers have been rotated to lie parallel (Fig. 1). For simplicity, we choose the torsional potential to have a minimum at $\phi=0$. Thus, neither ssDNA nor dsDNA will be helical in our model.

k_1 is chosen to be 0.1ϵ to give a persistence length, $l_{\mathrm{ps}}=3.149l$, at a reduced temperature of $T^*=0.096\,77$ for ssDNA. We obtain this result by simulating a single strand 70 bases in length, and using the definition[35]

$$l_{\mathrm{ps}} = \frac{\langle \mathbf{L}\cdot\mathbf{l}_1\rangle}{\langle l_1\rangle}, \tag{5}$$

where $\mathbf{L}$ is the end to end vector of the strand, $\mathbf{l}_1$ is the vector associated with the first monomer, and $\langle\ \rangle$ indicates a thermal average. Taking l to be 6.3 Å, $T^*=0.096\,77$ is mapped to 24 °C and so the model is consistent with experimental data for ssDNA in $0.445M$ NaCl solution.[36,37] k_2 is chosen to be 0.4ϵ.

In neglecting the geometrical structure of a double helix, we do not accurately represent certain types of bonding. "Bulged" bonding occurs when consecutive bases in one of the strands attach to nonconsecutive bases in the other strand. "Internal loops" consist of stretches of noncomplementary bases (either symmetric or asymmetric in the number of bases involved in each strand). "Hairpins" result when a single strand doubles back and bonds to itself. The details of these motifs are complicated but an empirical description of their thermodynamic properties is given in Ref. 26. Importantly, they are generally penalized due to the disruption of the geometry of DNA in a way that is not well reproduced by our model. In the case of the short strands we consider, these motifs will only play a small role as the strands are not specifically intended to have stable structures of these forms. In fact, as the base sequences we use were designed to form the Holliday junction, the possibility of forming these motifs at relevant temperatures was deliberately avoided.[38]

For simplicity, we therefore include only two alterations to the model. First, we define "kinked states" as those for which the number of unpaired bases between two bonding pairs on either side of a duplex is not equal (including asymmetric loops and bulges). We impose an infinite energy penalty on the formation of these kinked states if the total number of intermediate bases is less than 6. Second, we treat complementary units within six bases of each other on the same strand as noncomplementary but allow all other hairpins without penalty.

It should be also noted that this model neglects the directional asymmetry of the sugar-phosphate backbone. Therefore, parallel and antiparallel bonding are possible in our model, whereas parallel bonding does not occur in experiment.

B. Monte Carlo simulation

In a fully atomistic model of DNA, the natural way to simulate its dynamics would be to use molecular dynamics. However, the best way to simulate the dynamics in a coarse-grained model is an important, but not fully resolved, question, and one that will depend on the nature of the model. Clearly, for the current model, standard molecular dynamics is inappropriate as it will lead to ballistic motion of the strands between collisions because of the absence of explicit solvent particles, whereas DNA in solution undergoes diffusive Brownian motion. An alternative approach is to use the Metropolis Monte Carlo (MC) algorithm[39,40] where the moves are restricted to be local, as it has been argued that this can provide a reasonable approximation to the dynamics.[41–43] This is the approach that we use here to simulate the dynamics of self-assembly of DNA duplexes and Holliday junctions. In particular, the local MC moves that we use are the translation and rotation of whole strands and bending of a strand about a particular monomer, thus ensuring that the strands undergo an approximation to diffusive Brownian motion in the simulations. Therefore, we expect the MC simulations, which are all initiated with free single strands, to mimic the real self-assembly processes in our model. It is important to note that this will include, as well as successful assembly into the target structure, kinetic trapping in nonequilibrium configurations and that when the latter occurs this reflects the inefficiency of the self-assembly under those conditions.

Although a true measure of time is impossible in MC simulations, an approximate time scale for diffusion-limited processes can be found by comparing the diffusive properties

of objects to experiment. By measuring the diffusion of isolated strands, and assuming diffusion coefficients comparable to those of double strands and hairpin loops of similar length,[44] we conclude that one step per strand corresponds to a time scale of approximately 2 ps. Thus, our model allows our systems to be studied on millisecond time scales.

At the end of the above MC simulations, our systems will not necessarily have reached equilibrium, both because the energy barriers to escape from misbonded configurations can be difficult to overcome at low temperature and because of the low rate of association at higher temperatures. Therefore, as a comparison we also compute the equilibrium thermodynamic properties of our systems using umbrella sampling.[40,45] Formally, we can write the thermal average of a function $B(\mathbf{r}^N)$ in the canonical ensemble as

$$\langle B \rangle = \frac{\int \frac{B}{W(Q)}[W(Q)\exp(-V/k_BT)]d\mathbf{r}^N}{\int \frac{1}{W(Q)}[W(Q)\exp(-V/k_BT)]d\mathbf{r}^N}, \quad (6)$$

where $Q=Q(\mathbf{r}^N)$ is an order parameter or reaction coordinate and $V=V(\mathbf{r}^N)$ is the potential energy. We are free to choose $W(Q)$, and by taking the term in square brackets as the weighting of states and keeping statistics for B/W and $1/W$ at each step, we can find $\langle B \rangle$. In standard Metropolis MC, $W=1$, but by choosing $W(Q)$ in such a way that those states with intermediate values of Q are visited more frequently, the effective free energy barrier between (meta)stable states can be lowered allowing the system to pass easily between the free energy minima and equilibrium to be reached.

To ensure that each value of Q is equally likely to be sampled in an umbrella sampling simulation, one would choose $W(Q)=\exp(\beta A(Q))$, where $A(Q)$ is the free energy as a function of the order parameter. To achieve this, however, would require knowledge of $A(Q)$. Instead, there are standard methods to construct $W(Q)$ iteratively, but for the current examples it was possible to construct $W(Q)$ manually, because of the relative simplicity of the free energy profiles.

To a first approximation, the interaction between fully bonded structures is negligible. Therefore, in the umbrella sampling simulations, we consider systems containing the minimum number of strands required to form a given object (2 for a duplex and 4 for a Holliday junction). We then use the relative weight of bound and free states to extrapolate the expected fractional concentrations for larger systems.[46] The natural choice for the order parameter Q is the number of correct bonds, where two monomers are defined to be bonded if their energy of interaction is negative.

III. RESULTS

A. Duplex formation

We test the model by analyzing the duplex bonding of two different complementary strands. We simulate systems of ten oligonucleotides, initially not bonded, in a periodic cell with a concentration of 5.49×10^{-5} molecules l^{-3} (or $3.65\times10^{-4}M$). We separately consider strands consisting of 7 and 13 monomers, which correspond to two of the arms of the Holliday junction studied experimentally by Malo and co-workers[8,38] and which we consider in Sec. III B

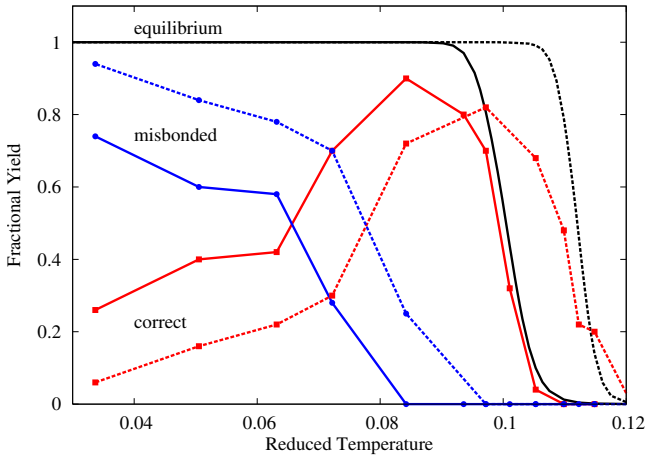

FIG. 2. (Color online) Yields of correctly formed duplexes and misbonded configurations at the end of our MC simulations (lines with data points, as labeled) compared to the equilibrium probability of the strands adopting the correct structure as obtained by umbrella sampling. The solid and dashed lines represent the results for strands with 7 and 13 monomers, respectively. The MC results are averages over ten runs of length 3×10^8 steps per strand with ten strands in the simulation cell.

$$7 \text{ bases} \begin{cases} \text{G-A-G-T-T-A-G} \\ \text{C-T-A-A-C-T-C}, \end{cases} \quad (7)$$

$$13 \text{ bases} \begin{cases} \text{G-C-G-A-T-G-A-G-C-A-G-G-A} \\ \text{T-C-C-T-G-C-T-C-A-T-C-G-C}, \end{cases} \quad (8)$$

where we have listed strands in the $5'-3'$ sense for consistency with the literature. The yields of correctly bonded and misbonded structures at the end of the simulations are depicted in Fig. 2 as a function of temperature. We also display the predicted equilibrium fraction of correctly bonded strands for both systems obtained using umbrella sampling. For convenience, we define a correctly bonded structure to have more than 70% of the bonds of the complete duplex and no bonds to other strands. Any other structure is recorded as "misbonded."

Figure 2 shows a maximum in the yield as a function of temperature. Such behavior is typical of self-assembling systems[47–51] and reflects the thermodynamic and dynamic constraints on the self-assembly process. First, the yield is zero at high temperature where only the ssDNA is stable, and rises just below the expected equilibrium value as the temperature is decreased, the deviation arising due to the large number of steps required to reach equilibrium. At low temperatures, the yield falls away due to the presence of kinetic traps which are now stable with respect to isolated strands, as evidenced by the rise in "misbonded structures" in Fig. 2. Thus, there is a nonmonotonic dependency of yield on temperature and an optimum region for successful assembly, which corresponds to the region where only the desired structure is stable against thermal fluctuations. Figure 3 is a snapshot from near the end of a simulation in this regime. It should be noted that neglecting helicity has the effect of increasing the flexibility of dsDNA in the direction perpen-

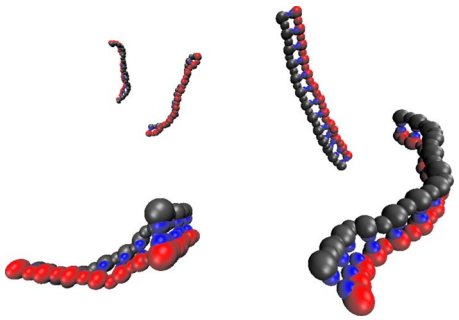

FIG. 3. (Color online) Snapshot of a fully assembled configuration in a MC simulation of ten 13-base strands at $T^*=0.0971$. Backbone sites are indicated by the large spheres, and bases by smaller spheres.

dicular to the plane of bonding (a helix cannot bend in any direction without disturbing its internal structure whereas a "ladder" can).

The heat capacity obtained from umbrella sampling of pair formation is shown in Fig. 4(a). The heat capacity peaks indicate a transition from single strands to a duplex. As the formation of duplexes is essentially a chemical equilibrium between monomers and clusters of a definite size (in this case 2), the width of the peaks will remain finite as the number of strands is increased. The transition does, however, become increasingly narrow as the DNA strands become longer, as is evident from comparing the heat capacity peaks for the 7-mer and 13-mers.

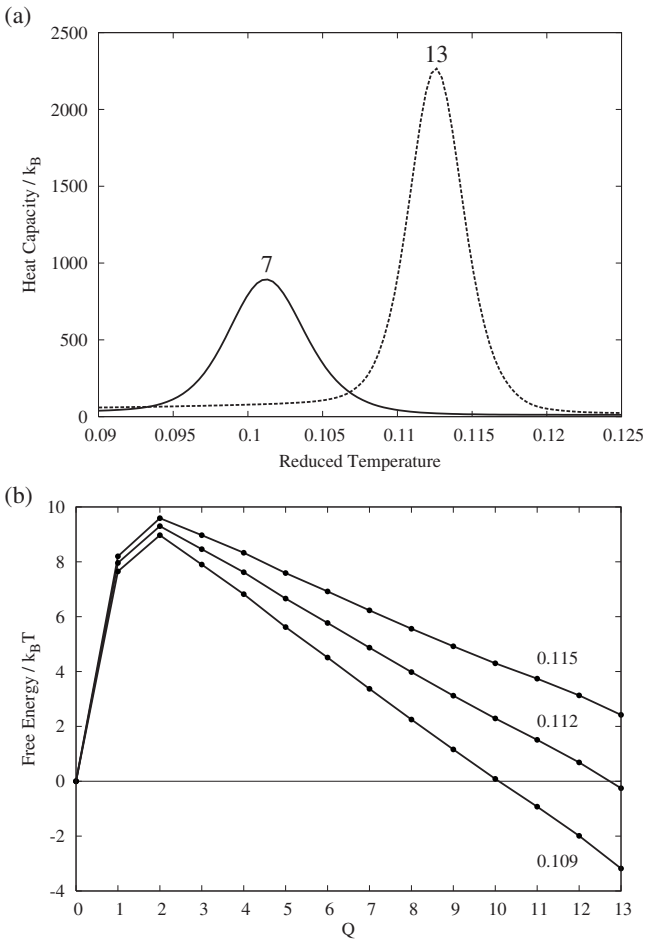

FIG. 4. Thermodynamics for the formation of a single duplex. (a) Heat capacity curves for 7- and 13-base systems, as labeled. (b) Free energy profile associated with the formation of a 13-base pair duplex at different temperatures (as labeled), where Q represents the number of correctly formed bonds in the duplex.

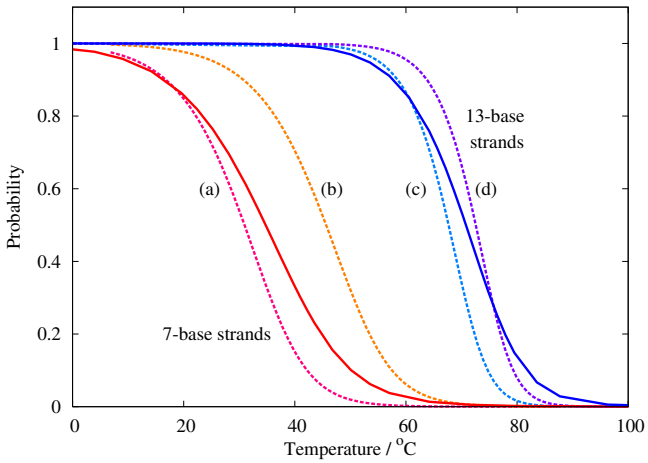

FIG. 5. (Color online) The bulk equilibrium probability of strands being in a correct duplex extrapolated from our umbrella sampling simulations (solid lines) compared to the predictions of an empirical two-state model (dashed lines). Results are presented for strands with 7 and 13 bases, as labeled. For the two-state model, as well as the results for the sequences in Eqs. (7) and (8) [lines (a) and (d)], sequences corresponding to the other two arms of the Holliday junction (Fig. 7) are considered. Only one line is shown for the umbrella sampling results, because A-T and G-C have the same binding energy in our model. Temperatures in our model are converted to °C using the mapping given in Sec. II A.

Figures 4(b) shows the free energy profile, $F(Q)$, for the formation of a duplex. The initial peak at low Q is accounted for by the entropic cost of bringing two strands together. Once bonds are formed, however, adding extra bonds costs much less entropy while providing a significant decrease in energy, explaining the monotonic decrease in $F(Q)$ beyond $Q=2$. The rise between $Q=1$ and $Q=2$ is partly due to the fact that in order to form two bonds between strands, the relative orientation of strands must be specified whereas this is not true for $Q=1$. Hence there is an additional entropy penalty to the formation of the second bond. In addition, there exist structures with only one correct bond that are stabilized by additional incorrect bonds and these misbonded configurations also contribute to $F(1)$. The constant gradient above $Q=2$ indicates that the energetic gain and entropic cost of forming an extra bond are approximately constant at a given temperature, which is consistent with the assumptions underlying nearest-neighbor models of DNA melting.[24]

In Fig. 5, we compare our melting curves to those predicted by a simple two-state model,[24] using the same mapping of the reduced temperature as in Sec. II A. In the two-state model, the molar concentrations of product (AB) and reactants (A and B) are given by the equilibrium relation

$$\frac{[AB]}{[A][B]} = \exp\left(\frac{-\Delta H_0 + T\Delta S_0}{k_B T}\right), \tag{9}$$

where ΔH_0 and ΔS_0 are assumed to be constants which depend only on the strand sequences and the salt concentration (we take $[Na^+]=0.445M$ as in Sec. II A). We use the enthalpy and entropy changes of duplex formation calculated

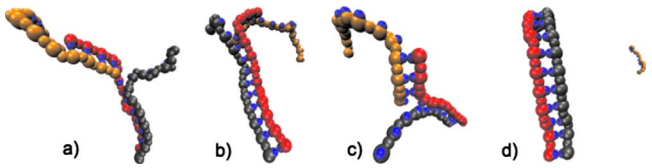

FIG. 6. (Color online) Snapshots illustrating four stages in the process of displacement. (a) A third strand binds to a misbonded pair. (b) The third strand is prevented from forming a complete duplex by the misbond. (c) Thermal fluctuations cause bonds in the misbonded structure to break and be replaced by the correct duplex. (d) The misbonded strand is displaced and the correct duplex is formed.

by "HYTHER," [52] a program that estimates these values using the "unified oligonucleotide nearest neighbor parameters." [25,53] The authors claim that the thermodynamic parameters predicted by HYTHER give the melting temperature T_m (the temperature at which the fraction of bonded strands is 1/2) of a duplex to within a standard error of $\pm 2.2\ ^\circ$C.[25] Figure 5 shows that our system reflects the melting temperatures predicted by the two-state model with reasonable accuracy, excepting sequence dependent effects which are not included in our model, because the interaction energies between A-T and C-G complementary base pairs have for simplicity been taken to be the same. The widths of the transitions are seen to be of the same order but slightly larger for our model. This feature, which is typical of coarse-grained models,[31] indicates that the degree of entropy loss on hybridization is too small in our model, and is due to a failure to accurately incorporate all degrees of freedom which become frozen on hybridization. However, the agreement is sufficiently good that the basic features of physical DNA assembly should be reproducible.

A further satisfying feature of the model is that "displacement" was observed on several occasions. This process, during which a misbonded pair of strands is broken up by a third strand, is illustrated in Fig. 6. The third strand is able to bond to the pair, as some bases are free in the misbonded structure. Thermal fluctuations allow the new strand to bond to sites previously involved in misbonding, in a process known as "branch migration." Eventually one of the misbonded strands is completely displaced, leaving a correct duplex and an isolated single strand. This behavior is observed in real DNA systems, and is the driving mechanism of some nanomachines[54] and DNA catalyzed reactions.[21,55]

B. Holliday junction

Encouraged by the above results, we next apply the model to the formation of a Holliday junction. Holliday junctions consist of four single strands which bind to form a four-armed cross. In our case we consider a Holliday junction with two long arms (13 bases long) and two short arms (7 bases long). We use the experimental base ordering of Malo *et al.*[38] with the "sticky ends" removed. (These sticky ends consist of six unpaired bases on the end of arms and their purpose is to allow the Holliday junctions to bond together to form a lattice.) The sequences of the four DNA strands and a schematic of the possible junctions that they can form are shown in Fig. 7.

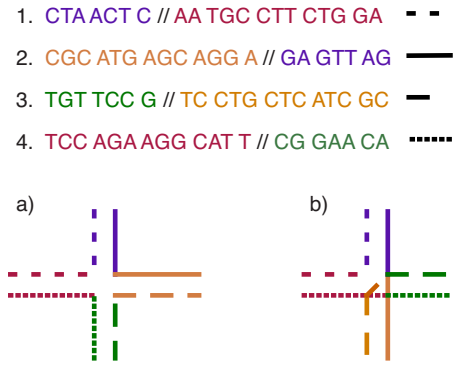

FIG. 7. (Color online) A schematic showing the sequences of the strands used in our Holliday junction simulations, and the alternative bound states that are possible: (a) The square planar configuration and (b) the χ-stacked form.

Initially we studied a system of 20 strands (five of each type) that has the potential to form five separate junctions. We use a concentration of 1.56×10^{-5} molecules l^{-3} (which corresponds to $1.04\times10^{-4}M$). The results are displayed in Fig. 8.

The results are as expected for the bonding of the longer arms (which we now describe as "α-bonding"). The yield again displays the characteristic nonmonotonic dependence on temperature. We obtain very few complete junctions, however, which is due to two effects. First, each simulation is performed at constant temperature, which means the hierarchical route to assembly is less favored than when the system is cooled, as in the experiments.[8,38] When the system is gradually cooled, Fig. 8 suggests that at around $T_m(\alpha)=0.111$ we would expect to find a region in which only α-bonded dimers were stable with respect to ssDNA. If the cooling was sufficiently slow on the time scale of bonding, all strands would form α-structures at around $T_m(\alpha)$. At

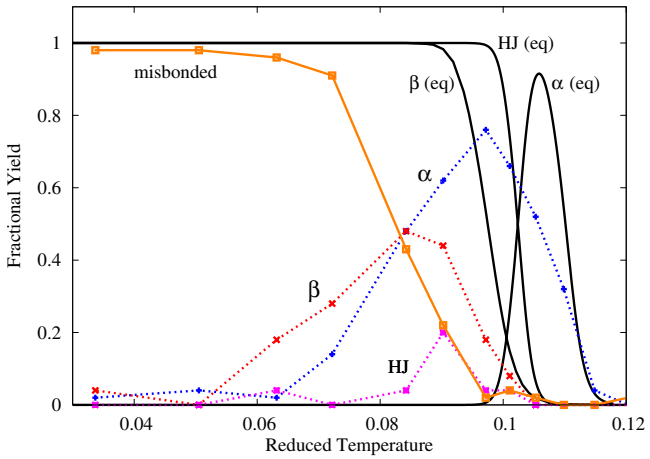

FIG. 8. (Color online) A comparison of the kinetics and thermodynamics for a system of 20 strands that can potentially form five Holliday junctions, where the MC simulations are initiated from a purely single-stranded configuration. The MC results (lines with data points) are the final yield of Holliday junctions, and the fraction of strands involved in a correctly formed long (α) or short (β) arm, or in misbonding, as labeled. The results are averages over five runs of length 10^9 steps per strand. For comparison, the equilibrium probabilities of being in an α-bonded dimer and a Holliday junction are also plotted, along with the equilibrium probability of being in a β-bonded dimer if the longer arms are not allowed to hybridize.

lower temperatures, when the Holliday junction becomes stable with respect to the α-bonded dimers, many competing minima would then be inaccessible to the system as they would require the disassociation of stable α-bonded pairs. The free energy landscape of two α-structures forming a Holliday junction is consequentially much simpler than that of four single strands forming a junction at a given temperature. Therefore, one expects the yield for self-assembly at constant temperature to be lower than when the system is cooled, because there is only a relatively narrow temperature window between where the Holliday junction becomes stable and misbonded configurations start to appear. Indeed, the shorter arms are only marginally more stable than some competing minima, as evidenced by the rise in misbonded structures in Fig. 8 at temperatures just below where β-bonded structures first appear.

Second, even in the temperature range where Holliday junction formation is not hindered by the formation of misbonded configurations, the yield is low because the Metropolis MC algorithm artificially reduces the diffusion of bound pairs, and hence the likelihood that two pairs of α-bonded strands come together to form a junction is also reduced. This is because the acceptance probability of trial moves for bonded strands is much lower than for isolated strands,[56] due to the energy penalty associated with trying to move a bound pair apart.

Interestingly, examination of the equilibrium lines in Fig. 8 shows that the Holliday junctions are actually stable at a higher temperature than the individual shorter arms. This is because the total loss of entropy when two α-bonded dimers bind together is considerably less than that for two short arms in isolation (as fewer translational degrees of freedom are lost), whereas the energy change is comparable. Thus, there is a small temperature window at $T^* \approx 0.1$ where hierarchical assembly can occur at constant temperature as the short arms are only stable once α-bonding has taken place. However, due to the deficiencies in the MC simulations mentioned above, the yield of Holliday junctions in this region is practically zero. Instead, the maximum yield of Holliday junctions occurs at lower temperatures where nonhierarchical pathways that proceed by the addition of single strands become feasible.

The above simulations were only able to successfully model the first stage of the Holliday junction assembly, namely, the formation of α-bonded dimers. To probe the second stage of assembly, we must first make two modifications to our simulation approach to overcome the two deficiencies mentioned above. First, we study systems initially consisting of pairs of α-bonded strands, which we assume have successfully formed at some higher temperature—this is reasonable given the results of our earlier simulations. Second, we also include simple local cluster moves in addition to those which move only one strand, i.e., translations, rotations, and bending of pairs of α-bonded strands. With these changes incorporated, we simulate the same system for 7.5×10^8 steps per strand at a range of temperatures below $T_m(\alpha)$. It should be noted that due to a change in the size of typical moves, one move per strand now corresponds to approximately 10 ps.

FIG. 9. (Color online) Snapshot showing five Holliday junctions formed at $T=0.0842$ after 5.67×10^8 MC steps per strand.

We find that Holliday junctions form over a wide range of intermediate temperatures, while kinetic traps at low temperature lead to incomplete bonding and consequently to the possibility of forming large clusters. A typical result from the high-yield regime is shown in Fig. 9. As fully bonded Holliday junctions are essentially inert, it is reasonable to analyze their assembly behavior by considering only one junction. The smaller system size has the effect of increasing the assembly rate because the strands have less distance to diffuse but does not affect the basic assembly mechanism. We therefore simulated systems consisting of two α-bonded pairs with the same concentration as above.

We also introduced some modifications to the umbrella sampling scheme in order to more efficiently compute the thermodynamics of the second stage of Holliday junction formation. As well as cluster moves, we also introduced a "tethering" component in the weighting function $W(Q)$. We introduce a length $r_{\min}$ that corresponds to the shortest distance between any pair of backbone sites on different strands. We then split $Q=0$ into two regions: We weight those states with $r_{\min} < 3l$ with $W=1$, but for $r_{\min} \geq 3l$ we use $W=0.1$. This enables us to increase the rate of transitions between $Q=0$ and 1, and reduces the time spent simply simulating the diffusion of α-bonded dimers waiting for a collision to occur.

The MC results are plotted in Fig. 10 along with the equilibrium results obtained from umbrella sampling. With the cluster moves in place, we now see a high yield of Holliday junctions and a broad maximum in the yield as a function of temperature. The hierarchical pathway has the effect suggested earlier. Namely, the temperature window over which correct formation can occur is vastly increased, as the most significant competing minima are inaccessible because their formation would require dissociation of the α-bonded pairs.

The model is therefore consistent with the experimentally observed hierarchical assembly of Holliday junctions as the system is cooled.[8,38] It should be noted, however, that the junctions in our model usually form in the "square planar" as opposed to the "χ-stacked" shape (Fig. 7) that is observed under normal experimental conditions. The preference for one structure is a subtle consequence of the concentration of cations and the precise helical geometry of DNA.[57] This level of detail is not included in our coarse-grained model, so

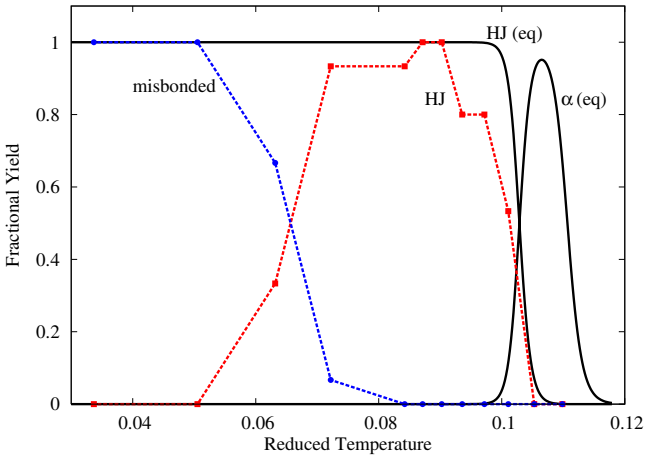

FIG. 10. (Color online) The yields of Holliday junctions and misbonded configurations for MC simulations, where the initial configuration was a pair of α-bonded dimers. For comparison, the equilibrium probabilities of being in an α-bonded dimer and a Holliday junction are also plotted. The results are averages over fifteen runs of length 7.5×10^8 steps per strand.

it is not surprising that it cannot reproduce the preference for χ-stacked structures. Moreover, it is relatively easy to see why in our model, which forms "ladders" rather than helices, a planar geometry is preferred for the junction.

Some of the equilibrium thermodynamic properties associated with the formation of a Holliday junction are shown in Fig. 11. In particular, Fig. 11(b) shows the free energy profile for the formation of a Holliday junction from two α-bonded pairs. The initial peak and subsequent drop is very similar to that for the duplexes and can be accounted for in the same way. However, the formation of the two arms is not like the zipping-up of a 14-base duplex, because there is much more relative freedom of movement for the bases on either side of the α-bonded sections in the dimers than for consecutive bases on single-stranded DNA. Thus, there is a rise between $M=7$ and $M=8$ that is a result of the entropy penalty of bringing together the two ends to make the second short arm. We note that the penalty is much smaller than the initial cost of bringing the two α-bonded pairs together, and as a result, the value of T_m for the junction is higher than for the short arms in isolation, as noted earlier.

An interesting feature of Fig. 11(b) is the plateau between $Q=6$ and $Q=7$. In general, when two α-bonded pairs meet to form one short arm, there is an entropic penalty associated with the excluded volume that the remaining bases in the α-structures represent to each other. This excluded volume is a large fraction of the total available space if one complete short arm is formed, so that there are no free monomers between the short arm and the α-bonded sections. As a consequence, there is not the usual free energy benefit from forming the final bond in the short arm (the one closest to the center of the Holliday junction), as the excluded volume penalty is large and those states that are allowed involve distortion of the backbones and bonds near the center of the Holliday junction. Although the details of this free energy penalty and the other features in Fig. 11(b) will depend on the exact geometry of the system, we expect the calculated free energy profile to be representative of that for real DNA.

It is possible to extend the two-state model discussed in

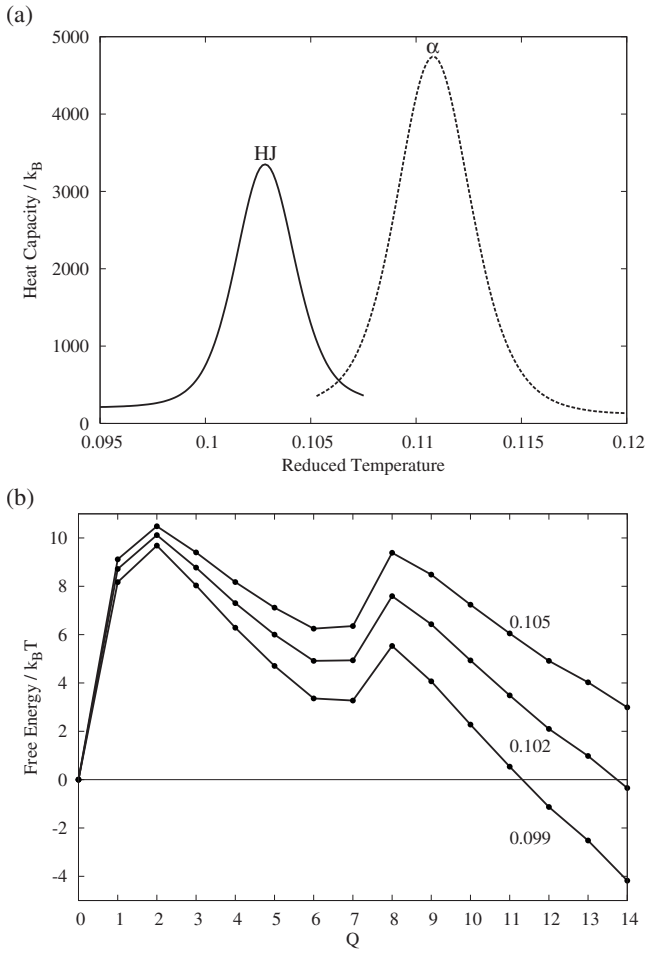

FIG. 11. Thermodynamics for the formation of a single Holliday junction. (a) Heat capacity curves for two α-bonded dimers forming a Holliday junction and four strands forming two α-bonded dimers, as labeled. (b) Free energy profiles for the formation of a Holliday junction from two α-bonded dimers at different temperatures, as labeled. Q represents the total number of correct bonds in the short arms of the junction.

Section III A to the formation of a Holliday junction by considering the concentrations of all four isolated strands, the two α-bonded intermediates and the junction itself. We assume that Eq. (9) holds for every possible transition, use the same thermodynamic parameters as before, and apply conservation of total strand number. To estimate ΔH_0 and ΔS_0 associated with the formation of a Holliday junction, we construct a single strand by linking the ends of the oligonucleotides together with four nonbonding bases. The thermodynamic parameters associated with the folding of this structure are predicted by "UNAFold." [58] The correction for the fact that our strands are not connected by loops is discussed by Zuker.[59] This leaves five simultaneous equations (assuming perfect stoichiometry) which can be solved numerically.

Figure 12 compares this extended two-state model (ETSM) with the bulk thermodynamics predicted by umbrella sampling (using the same temperature scaling as before). ETSM predictions for both stages of Holliday junction formation agree well with our results, which again supports our hypothesis that much of the physics of self-assembly can be reproduced by a simple coarse-grained model. The extra

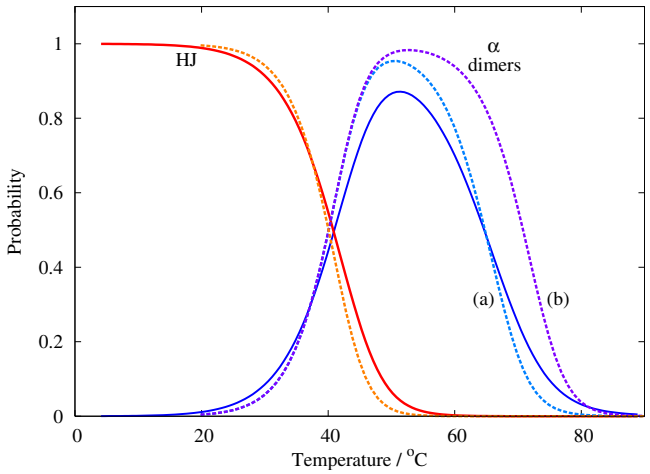

FIG. 12. (Color online) Bulk equilibrium probability of strands being in a Holliday junction or an α-bonded dimer computed by umbrella sampling (solid lines) and by the ETSM (dashed lines), where lines (a) and (b) represent the two possible α-bonded dimers.

width of the transitions in our model occurs for the same reasons as mentioned in Sec. III A when discussing Fig. 5.

C. Negative design

The hierarchical pathway for the formation of the Holliday junction is one aspect of the sequence design that aids the formation of the correct structure. The experimental base ordering of the Holliday junction, however, was also chosen to minimize the number of competing structures—a typical example of "negative design." [60] We illustrate the importance of such negative design by considering a badly designed junction, where the complementary seven-base sections consist of just one base type each. We simulate a system of two α-bonded pairs with the short arms so modified under the same conditions as for Fig. 10 and the results are shown in Fig. 13. Although there is some probability of forming the correct junction, the simulations are dominated by misbonded junctions, such as the one depicted in Fig. 14. Although these competing structures are energetically less stable than the target junction because of the presence of unpaired bases at the "dangling" ends, they are readily accessible because the likelihood that the first bonds formed between two α-bonded pairs are in the same registry as the target structure is low. The yield of the correct Holliday junctions will then depend on how readily the system is able escape from these malformed junctions. Clearly, this process is slow on the time scales of the current simulations, and is also likely to hinder the location of the target structure in experiment.

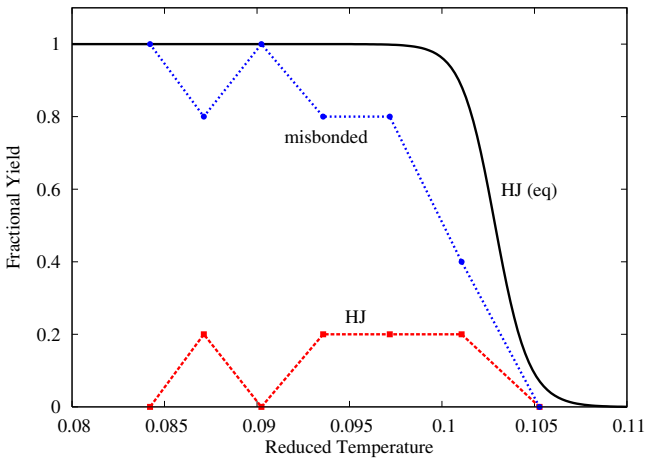

FIG. 13. (Color online) Simulation results for the badly designed Holliday junction of Sec. III C, where the initial configuration was a pair of α-bonded dimers. The lines with data points give the yield of correctly formed junctions and misbonded configurations, as labeled. For comparison, the solid line gives the equilibrium probability that the well-designed sequences of Section III B adopt a Holliday junction.

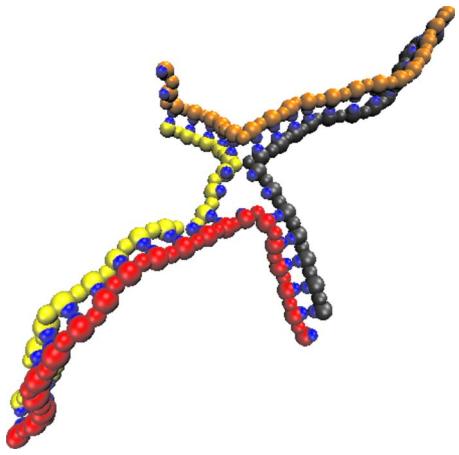

FIG. 14. (Color online) Example of a competing minimum for a badly designed Holliday junction. The snapshot is taken from a simulation at $T^*=0.0936$.

IV. DISCUSSION

In this paper, we have introduced a simple coarse-grained model of DNA in order to test the feasibility of modeling the self-assembly of DNA nanostructure by MC simulations. Any such model involves a trade-off between detail and computational simplicity, and here we deliberately chose to keep the model as simple as possible in order to give us the best chance of being able to probe the time scales relevant to self-assembly. The model involves just two interaction sites per nucleotide.

The results from our model are very encouraging. First, we have shown that using our model it is feasible to model the self-assembly of both DNA duplexes and a Holliday junction. The latter represents, to the best of our knowledge, the first example of the simulation of the self-assembly of a DNA structure beyond a duplex. Second, the model succeeds in reproducing many of the known thermodynamic and dynamic features of this self-assembly. For example, the equilibrium melting curves agree well with those predicted by the nearest-neighbor two-state model,[25] which is known to predict melting temperatures very accurately. The model is also able to capture important dynamical phenomena such as displacement.

Third, by analyzing the thermodynamic and dynamic constraints on assembly, we have been able to gain some important physical insights into the nature of DNA self-assembly and how to control it. For example, the optimal conditions for self-assembly are in the temperature range just

below the melting temperature of the target structure, where this structure is the only one stable with respect to the precursors, be they ssDNA or some intermediate in a hierarchical assembly pathway. At lower temperatures, misbonded configurations can be formed that act as kinetic traps and reduce the assembly yield. Similar trade-offs between the thermodynamic driving force and the kinetic accessibility have been previously seen in a variety of self-assembling systems,[47–51] and also give rise to a maximum in the yield near to and below the temperature at which the target structure becomes stable.

We have also seen how hierarchical self-assembly through cooling can be a particularly useful strategy to aid self-assembly, because the formation of stable intermediates at higher temperatures simplifies the free energy landscape for the assembly of the next stage in the hierarchy by reducing the number of misbonded configurations available to the system. This simplification of the energy landscape is likely to be a general feature of hierarchical self-assembly.

Thus, our results have confirmed the utility of using coarse-grained DNA models to study the self-assembly of DNA nanostructures, and supported our hypothesis that much of the physics can be explained by describing DNA as a semiflexible polymer with selective attractive interactions. The model's success in forming junctions in reasonable computational time suggests that it will be possible to develop further models that have an increased level of detail but which can still access the time scales relevant to self-assembly.

The model has also highlighted some features which it would be advantageous to include in such models. For example, greater accuracy in the details of oligonucleotide geometry, particularly the helicity of dsDNA, would allow features such as the characteristically long persistence length of hybridized strands to be reproduced and give the appropriate degree of rigidity to simulated nanostructures. Such improvement might also allow more complicated motifs to be accounted for, such as the preference for χ-stacked Holliday junctions that the current model could not reproduce.

It should be noted that if one is to introduce helicity in a physically reasonable way, it should also allow for ssDNA to undergo a stacking transition to a helical form. This transition may play a significant role in the thermodynamics and kinetics of self-assembly.[61] Previously proposed coarse-grained DNA models that incorporate helicity have not been designed to accurately reproduce this feature. Incorporating extra degrees of freedom which are relevant to the stacking transition, such as the rotation of the base with respect to the sugar-base bond, may also help to increase the entropy change on hybridization and hence make the transition narrower as required.

The approximation to diffusive dynamics provided by the local move Metropolis MC algorithm could also be improved. Currently the "local" moves involve displacing, rotating, or bending entire strands or pairs of strands—these effectively constitute cluster moves of groups of strongly bound nucleotides and result in slow relaxation and translation times within bound structures. More realistic dynamics may be achievable by considering trial moves of individual nucleotides and incorporating cluster moves in a more systematic fashion, such as in the "virtual move" MC algorithm proposed by Whitelam and Geissler.[51,62]

One potential issue with any coarse graining is how it preserves the different time scales in a system. In Sec. II B, we assigned an approximate mapping between the number of MC steps and physical time based on comparison of diffusion coefficients. There are, however, other important time scales in the system, such as the time scale for the internal dynamics of an isolated strand and the time scale over which the "zipping-up" of two strands occurs after a bond has been formed. Comparisons of experimental diffusion coefficients[44] and melting and bubble formation from molecular dynamics simulations[30,31] suggest a large separation in time scale between diffusion-limited processes and those that rely on the dynamics of individual nucleotides. Encouragingly, we observe a similar time scale separation in our model: Zipping-up and thermal relaxation of isolated strands occur over time scales shorter than 10^5 steps per strand, whereas association typically required on the order of 10^7-10^8 steps per strand near the melting temperature (corresponding to tens or hundreds of microseconds). Furthermore, we would argue that it is this time scale separation, and not the precise ratios of the relevant rate constants, that is important to reproduce in self-assembly simulations.

We should also note that the mapping of the diffusion constants between the model and experiment will not necessarily ensure that the rate of association is accurate in our model, because although the frequency of collisions in our model should be correct, there is also the contribution to the association rate from the probability that a collision will lead to successful association. That we can reproduce the thermodynamics of the DNA melting transitions implies that the rates of association and disassociation have the right ratio but not that they necessarily have the correct absolute value. For example, it is conceivable that helicity (both in dsDNA and possibly in ssDNA), which is not included in the current model, will influence the likelihood that a collision is successful.

ACKNOWLEDGMENTS

The authors are grateful for financial support from the EPSRC and the Royal Society. They also wish to acknowledge helpful discussions with Jonathan Malo, John Santalucia Jr., and Michael Zuker.

[1] N. C. Seeman, Nature (London) **421**, 427 (2003).
[2] S. Pitchiaya and Y. Krishnan, Chem. Soc. Rev. **35**, 1111 (2006).
[3] W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, Berlin, 1984).
[4] P. J. Hagerman, Annu. Rev. Biophys. Biophys. Chem. **17**, 265 (1988).
[5] N. R. Kallenbach, R.-I. Ma, and N. C. Seeman, Nature (London) **305**, 829 (1983).
[6] T. J. Fu and N. C. Seeman, Biochemistry **32**, 3211 (1993).
[7] E. Winfree, F. R. Liu, L. A. Wenzler, and N. C. Seeman, Nature (London) **394**, 539 (1998).
[8] J. Malo, J. C. Mitchell, C. Venien-Bryan, J. R. Harris, H. Wille, D. J. Sherrat, and A. J. Turberfield, Angew. Chem., Int. Ed. **44**, 3057 (2005).
[9] H. Yan, S. H. Park, G. Finkelstein, J. H. Reif, and T. H. LaBean, Science **301**, 1882 (2003).
[10] P. W. K. Rothemund, Nature (London) **440**, 297 (2006).
[11] J. H. Chen and N. C. Seeman, Nature (London) **350**, 631 (1991).

[12] Y. W. Zhang and N. C. Seeman, J. Am. Chem. Soc. **116**, 1661 (1994).
[13] R. P. Goodman, I. A. T. Sharp, C. F. Tardin, C. M. Erben, R. M. Berry, C. F. Schmidt, and A. J. Turberfield, Science **310**, 1661 (2005).
[14] C. M. Erben, R. P. Goodman, and A. J. Turberfield, J. Am. Chem. Soc. **129**, 6992 (2007).
[15] W. M. Shih, J. D. Quispe, and G. F. Joyce, Nature (London) **427**, 618 (2004).
[16] F. F. Andersen, B. Knudsen, C. L. P. Oliveira, R. F. Frøhlich, D. Krüger, J. Bungert, M. Agbandje-McKenna, R. McKenna, S. Juul, C. Veigaard, J. Koch, J. L. Rubinstein, B. Gulbrandtsen, M. S. Hede, G. Karlsson, A. H. Andersen, J. S. Pedersen, and B. R. Knudsen, Nucleic Acids Res. **36**, 1113 (2007).
[17] Y. He, T. Ye, M. Su, C. Zhang, A. Ribbe, W. Jiang, and C. Mao, Nature (London) **452**, 198 (2008).
[18] F. A. Aldaye and H. F. Sleiman, J. Am. Chem. Soc. **129**, 13376 (2007).
[19] J. Zimmermann, M. P. J. Cebulla, S. Mönninghoff, and G. von Kiedrowski, Angew. Chem., Int. Ed. **47**, 3626 (2008).
[20] C. Pistol, A. R. Lebeck, and C. Dwyer, in *DAC '06: Proceedings of the 43rd Annual Conference on Design Automation* (ACM, New York, 2006), pp. 919–924.
[21] P. Yin, H. M. Choi, C. R. Calvert, and N. A. Pierce, Nature (London) **451**, 318 (2008).
[22] T. E. Cheatham III, Curr. Opin. Struct. Biol. **14**, 360 (2004).
[23] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers: Statistical Mechanical Theory of Order-Disorder Transitions in Biological Macromolecules* (Academic, New York, 1970).
[24] R. Everaers, S. Kumar, and C. Simm, Phys. Rev. E **75**, 041918 (2007).
[25] J. SantaLucia, Jr., Proc. Natl. Acad. Sci. U.S.A. **95**, 1460 (1998).
[26] J. SantaLucia, Jr. and D. Hicks, Annu. Rev. Biophys. Biomol. Struct. **33**, 415 (2004).
[27] J. Bois, S. Venkataraman, H. M. T. Choi, A. J. Spakowitz, Z. Wang, and N. A. Pierce, Nucleic Acids Res. **33**, 4090 (2005).
[28] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, 684 (1993).
[29] S. Buyukdagli, M. Sanrey, and M. Joyeux, Chem. Phys. Lett. **419**, 434 (2006).
[30] K. Drukker, G. Wu, and G. C. Schatz, J. Chem. Phys. **114**, 579 (2001).
[31] T. A. Knotts IV, N. Rathore, D. Schwartz, and J. de Pablo, J. Chem. Phys. **126**, 084901 (2007).
[32] J. C. Araque, A. Z. Panagiotopoulos, and M. A. Robert, (unpublished).
[33] M. Sales-Pardo, R. Guimera, A. A. Moreira, J. Widom, and L. Amaral, Phys. Rev. E **71**, 051902 (2005).
[34] F. W. Starr and F. Sciortino, J. Phys.: Condens. Matter **18**, L347 (2006).
[35] P. Cifra, Polymer **45**, 5995 (2004).
[36] M. C. Murphy, I. Resnik, W. Chang, T. M. Lohman, and T. Ha, Biophys. J. **86**, 2530 (2004).
[37] As with all "ladder" models, there is an ambiguity in defining the value of l because it is also equal to the separation of bases along the chain in the double-stranded geometry, whereas in the double helix the rise per base pair is 3.3 Å due to the wrapping of single strands around the axis. As the absolute value of l is used to map onto the experimental persistence length of ssDNA, we choose to use 6.3 Å.
[38] J. Malo, Ph.D. thesis, Oxford University, 2005.
[39] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[40] D. Frenkel and B. Smit, *Computer Simulation of Liquids* (Academic, London, 2001).
[41] K. Kikuchi, M. Yoshida, T. Maekawa, and H. Watanabe, Chem. Phys. Lett. **185**, 335 (1991).
[42] L. Berthier and W. Kob, J. Phys.: Condens. Matter **19**, 205130 (2007).
[43] G. Tiana, L. Sutto, and R. A. Broglia, Physica A **380**, 241 (2007).
[44] J. Lapham, J. P. Rife, P. B. Moore, and D. M. Crothers, J. Biomol. NMR **10**, 255 (1997).
[45] G. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[46] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye (unpublished).
[47] H. D. Nguyen, V. S. Reddy, and C. L. Brooks III, Nano Lett. **7**, 338 (2007).
[48] M. F. Hagan and D. Chandler, Biophys. J. **91**, 42 (2006).
[49] A. W. Wilber, J. P. K. Doye, A. A. Louis, E. G. Noya, M. A. Miller, and P. Wong, J. Chem. Phys. **127**, 085106 (2007).
[50] D. C. Rapaport, Phys. Rev. Lett. **101**, 186101 (2008).
[51] S. Whitelam, E. H. Feng, M. F. Hagan, and P. L. Geissler, "The role of collective motion in examples of coarsening and self-assembly," Soft Matter (in press); arXiv:0806.390.
[52] N. Peyret and J. SantaLucia, Jr., HYTHER™ version 1.0, URL: http://ozone3.chem.wayne.edu/cgi-bin/login/.
[53] N. Peyret, P. A. Seneviratne, H. T. Allawi, and J. SantaLucia, Jr., Biochemistry **38**, 3468 (1999).
[54] B. Yurke, A. J. Turberfield, A. P. Mills, Jr., F. C. Simmel, and J. Neumann, Nature (London) **406**, 605 (2000).
[55] D. Y. Zhang, A. Turberfield, B. Yurke, and E. Winfree, Science **318**, 1121 (2007).
[56] E. Luijten, Comput. Sci. Eng. **8**, 20 (2006).
[57] M. Ortiz-Lombardia, A. Gonzalez, R. Eritja, J. Aymami, F. Azorin, and M. Coll, Nat. Struct. Biol. **6**, 913 (1999).
[58] N. R. Markham and M. Zuker, in *Bioinformatics: Volume II: Structure, Functions and Applications*, edited by J. M. Keith (Humana, Totowa, 2008).
[59] M. Zuker, Nucleic Acids Res. **31**, 3406 (2003).
[60] J. P. K. Doye, A. A. Louis, and M. Vendruscolo, Phys. Biol. **1**, P9 (2004).
[61] J. Holbrook, M. Capp, R. Saecker, and M. Record, Biochemistry **38**, 8409 (1999).
[62] S. Whitelam and P. L. Geissler, J. Chem. Phys. **127**, 154101 (2007).