DNA Nanotweezers Studied with a Coarse-Grained Model of DNA

Thomas E. Ouldridge,[1] Ard A. Louis,[1] and Jonathan P. K. Doye[2]

[1]*Rudolf Peierls Centre for Theoretical Physics, 1 Keble Road, Oxford, OX1 3NP, United Kingdom*
[2]*Physical and Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford, OX1 3QZ, United Kingdom*

(Received 3 November 2009; published 26 April 2010)

We introduce a coarse-grained rigid nucleotide model of DNA that reproduces the basic thermodynamics of short strands, duplex hybridization, single-stranded stacking, and hairpin formation, and also captures the essential structural properties of DNA: the helical pitch, persistence length, and torsional stiffness of double-stranded molecules, as well as the comparative flexibility of unstacked single strands. We apply the model to calculate the detailed free-energy landscape of one full cycle of DNA ''tweezers,'' a simple machine driven by hybridization and strand displacement.

DOI: 10.1103/PhysRevLett.104.178101 PACS numbers: 87.14.gk, 34.20.Gj, 81.07.Nb, 87.15.A−

The field of DNA nanotechnology has grown rapidly in recent years as investigators have harnessed the selectivity of DNA base pairing to form many different kinds of structures. Recent examples include large ribbons [1], two-dimensional lattices [2], and polyhedra [3,4], made by hybridizing systems of short strands (oligonucleotides). Another technique, DNA origami [5], uses short ''staple'' strands to fold a long polynucleotide into almost any two-dimensional shape, and has recently been extended to three-dimensional structures [6,7].

Hybridization free energy can also be harnessed in artificial molecular machines. Simple designs, such as DNA ''tweezers'' [8], use alternating addition of strands to drive a system through a conformational cycle. More sophisticated, autonomous machines catalyze the hybridization of strands initially present in inert forms, such as hairpins. Hybridization cycles can be coupled to directional motion on a DNA track, creating DNA walkers [9,10].

Computer simulations of these DNA nanosystems would complement experiments with detailed insight into the processes involved in assembly or mechanical cycles. The system sizes and time scales involved make all-atom simulations prohibitively expensive: instead, coarse-grained models must be used. A common method is to integrate out the microscopic degrees of freedom from an atomistic model. Instead, we take a ''top-down'' approach that focuses on capturing the generic properties of DNA that are important for self-assembly. In particular, we choose base-pairing and stacking interactions to reproduce the structural and thermodynamic changes associated with duplex formation. Consequentially, the model can simultaneously reproduce duplex hybridization, single-stranded stacking, and hairpin formation, as well as important physical properties such as the helical pitch, persistence length, and torsional stiffness of double-stranded molecules, and the comparative flexibility of unstacked single strands. This approach makes detailed studies of DNA nanostructures possible, which we show by simulating a full cycle for DNA tweezers.

A number of other coarse-grained models of DNA have been suggested in recent years. Nonhelical models, with two interaction sites per nucleotide, have been applied to duplex [11], hairpin [12,13], and four-arm junction formation [11], as well as the gelation of oligonucleotide functionalized colloids [14]. Helical models with two [15] or three [16,17] sites per nucleotide have been used to study denaturation and hybridization of double-stranded DNA. However, to study the formation of nanostructures or the operation of hybridization-driven nanodevices, it is essential to have a physically reasonable representation of both single- and double-stranded states. Earlier models either neglect the helicity of double-stranded DNA, or impose it through restrictions on the backbone of a single strand, which leads to an unphysical representation of single-stranded DNA. Furthermore, the thermodynamic properties of hybridization (particularly the widths of transitions) have not been well reproduced.

We model DNA as a string of rigid nucleotides, as depicted in Fig. 1, with one interaction site for the backbone and three for the base. An additional ''normal vector'' indicates the plane of the base. Backbone sites are connected via finitely extensible nonlinear elastic (FENE) springs, and act as soft repulsion centers (along with the base repulsion sites) to reproduce steric interactions.

The helicity of our model results directly from the stacking interactions between base stacking interaction

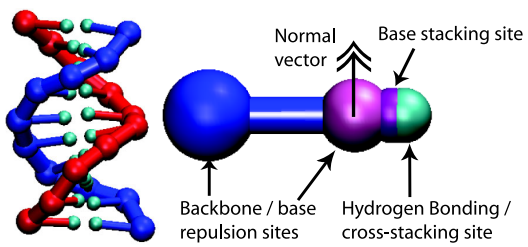

FIG. 1 (color online). A duplex as represented by the model and a detailed view of a nucleotide.

sites. Consecutive bases attract each other with a minimum at approximately 3.4 Å, shorter than the equilibrium FENE spring length of approximately 6.5 Å. We modulate this interaction according to the relative alignment of the normal vectors, and the alignment of the normals with the intersite vector. This drives the formation of helical stacks of coplanar bases: right handedness results from setting the attraction to zero if bases stack left-handedly.

Hydrogen bonding is represented by an attraction between hydrogen bonding sites of complementary bases, modulated by factors favoring colinear nucleotides with antiparallel normals. With the stacking interaction, hydrogen bonding drives the formation of right-handed double helices with the approximate geometry of B-DNA. We also include a cross-stacking interaction between bases that are diagonally opposite each other in a double helix, enabling the tendency of ''dangling ends'' to stabilize duplexes to be reproduced. The complete form of all potentials can be found in the Ref. [18].

For simplicity, several features of DNA have been neglected. First, although only complementary bases can bond in our model, all bases are otherwise identical; at this stage, we are interested in the generic properties of DNA assembly rather than specific base heterogeneity effects. Secondly, we fit the parameters using experimental data at just one salt concentration, $[\mathrm{Na}] = 500$ mM, where the Debye screening length is short ($\sim$4.5 Å) and most properties are only weakly salt dependent. Finally, major and minor grooving are neglected.

We simulate the model using the ''virtual move Monte Carlo'' algorithm of Whitelam and Geissler [19]. Because of the system's simplicity and the efficiency of the algorithm, denaturation and hybridization of short duplexes occur without biasing the ensemble. For accurate statistics, however, we use umbrella sampling [20] to characterize the basic DNA transitions.

ssDNA undergoes a transition from an ordered, helical form at low temperature to a disordered structure at high temperature [21]. Our model reproduces a broad, almost uncooperative transition with an enthalpy of $\Delta H^{\mathrm{stack}} = -5.6$ kcal mol^{-1} and entropy of $\Delta S^{\mathrm{stack}} = -16.5$ cal mol^{-1} K^{-1}, similar to experimental estimates from Holbrook *et al.* [22] of $\Delta H^{\mathrm{stack}} = -5.7$ and -5.3 kcal mol^{-1} and $\Delta S^{\mathrm{stack}} = -16.0$ and -15.0 cal mol^{-1} K^{-1} for two strands. Although these values were found at $[\mathrm{Na}^+] = 120$ mM, similar results at 50 mM suggest weak salt dependence in this regime.

We study duplex formation by simulating two complementary strands in a box at an effective concentration of 0.317 mM, extrapolating to bulk using the method in Ref. [23]. We compare to melting temperatures (T_m) predicted by the nearest-neighbor model of SantaLucia [24], for strands consisting of ''average bases'' [18] (with thermodynamic parameters averaged over all complementary base pair steps). Figure 2 shows that our model agrees with the predictions for T_m over a range of duplex lengths.

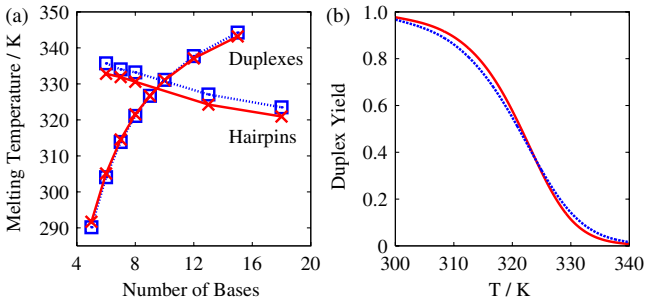

FIG. 2 (color online). (a) Comparison of melting temperatures as computed for our model (crosses connected by a solid line) and predicted by the nearest-neighbor model [24] (squares connected by a dashed line) for duplexes as a function of the single-stranded length, and hairpins as a function of loop length for a stem of six bases. (b) Melting profile for an eight base duplex as predicted by our model (solid line) and the nearest-neighbor model (dashed line).

Importantly, transition widths also coincide to within approximately 2 K, and thus the agreement in T_m will hold over a range of concentrations.

The third basic transition is hairpin formation, in which self-complementary strands bind to themselves to form a stem and hairpin loop. Our model underestimates T_m relative to the nearest-neighbor model by approximately 3 K (less than 1% of the absolute temperature), but importantly captures the dependence on loop (Fig. 2) and stem length.

In addition to thermodynamics, the model reproduces many of the physical properties of DNA essential for nanotechnology. Model duplexes have a pitch of 10.4 base pairs per turn, a persistence length of 154 base pairs, and an RMSD of 3.7° in the twist of each base pair rise. Unstacked single strands are comparatively flexible, having a persistence length of 19.5 Å (we define model length scales so that the average rise per base pair at 300 K is 3.3 Å). These values compare favorably with reported experimental results of 10.5 base pairs per turn [25], 135–150 base pairs [26], 3.9° [26], and 19.4 Å [27], respectively.

Having demonstrated that our model reproduces the essential physics of DNA assembly, we apply it to ''DNA tweezers,'' a simple example of DNA hybridization driving conformational changes [8]. The cycle is shown in Fig. 3, with the tweezer unit switching between open and closed conformations as fuel (f) and antifuel ($\bar{f}$) strands are sequentially added, producing an $f\bar{f}$ duplex as waste.

For simplicity, we simulate a system approximately half the size of that originally used by Yurke *et al.* [8], with the sequences listed in Ref. [18]. The tweezers themselves consist of three strands [a hinge strand and two arms (α and β)], forming two duplex regions of ten base pairs connected by a flexible, single-stranded hinge of four bases. At the end of the duplexes, there are overhanging single-stranded sections of eight bases. The fuel f is 24 bases in length, and is complementary to the overhanging regions of the tweezers, enabling it to bind to both and close the tweezers [Fig. 3(c)]. The additional

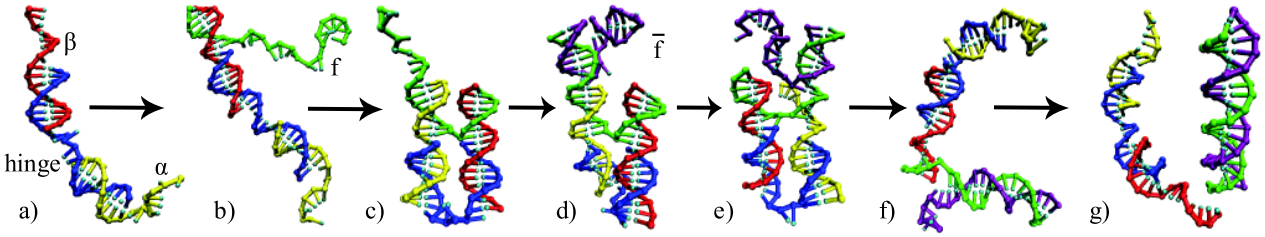

FIG. 3 (color online). Simulation snapshots showing stages of operation of DNA tweezers. (a) Tweezers initially open. (b) Fuel (f) is added and binds to one arm (β). (c) Fuel binds to the second arm (α) and closes the tweezers. (d) Antifuel ($\bar{f}$) is added and binds to the toehold of the fuel. (e) Antifuel begins to displace first arm of the tweezers. (f) Tweezers open as first arm is displaced, and antifuel starts to displace the second arm. (g) Antifuel fully hybridizes to fuel and the waste duplex is formed.

eight bases provide a "toehold" for binding of the antifuel $\bar{f}$, which is also 24 bases long and complementary to the whole of f.

The tweezers, like many DNA based machines, rely on toehold-mediated strand displacement [28]. After the addition of $\bar{f}$, the closed structure becomes metastable as the free energetic minimum of the system is an $f\bar{f}$ duplex isolated from the tweezers. $\bar{f}$ can bind to the toehold of f [Fig. 3(d)]: $\bar{f}$ and α then compete for binding to the rest of f. By binding to available bases, $\bar{f}$ reduces the free-energy barrier for dissociation of f from α, thereby accelerating the approach to equilibrium. Once α is displaced, the process is repeated with β.

We have sampled the free-energy landscape of the system consisting of one set of tweezers and a single f and $\bar{f}$, in a periodic cell of volume 4.19×10^5 nm^3 (Fig. 4). Every stage of the cycle is observable using unbiased simulations at 300 K. To obtain the free-energy landscape, however, we split the order parameter space into umbrella sampling windows, which were then combined using the weighted histogram analysis method [29]. Further details on how the sampling was performed are given in Ref. [18].

To study the cycle in detail, it is convenient to consider a one-dimensional pathway through the landscape; we use that shown by the arrows in Fig. 4. The free-energy difference between "a" and "g" is found to be approximately 47 kT; simulations of f and $\bar{f}$ in isolation (displayed along $y = 0$ in Fig. 4) gave a free-energy difference which was consistent to within statistical error [18].

The gross features of the free-energy landscape are as expected. Duplex formation is highly cooperative; pairing two strands involves a high entropic cost for forming the first base pair, then a downhill slope in free energy as additional bonds form [17]. This is reflected in Fig. 5 by stages "b," "c," and "d" which essentially involve duplex formation. This cooperativity suggests that f will fully bind to one arm of the tweezers before binding to the second. The displacement processes (indicated by "e" and "f" in Fig. 4) are comparatively flat as the total number of interstrand base pairs is constant. Returning the tweezers to the open state (between "e" and "f") and the decoupling of the $f\bar{f}$ duplex from the tweezers ("g") release the free energy stored in bringing strands together, resulting in large decreases in free energy.

Simulations allow for a detailed inspection of processes like displacement. Thus, Fig. 5 shows that there is actually an increase in free energy of ~3 kT during the displacement of the first strand α, even though the total number of interstrand base pairs in the system stays constant. The increase in free energy with displacement is initially steady, with a sharper jump after four bases, followed by another smooth increase. Conversely, the displacement of the second strand β shows a steady decrease in free energy as more bases are displaced. These slopes suggest a significant difference in speed for the two processes: our unbiased simulations show that the first displacement requires about 10 times as many Monte Carlo moves, suggesting a slow displacement of the first arm, followed by a quicker displacement of the second.

Two effects help to explain the increase in free energy during the displacement of α. First, $\bar{f}$ is capable of forming a hairpin structure, as shown in Fig. 3(d), which is marginally stable at 300 K. After the displacement of four bases of α, however, the hairpin can no longer form, leading to the observed step up in free energy. Simulations were performed in which the final eight bases of $\bar{f}$ were prevented from forming hairpins [Fig. 5(b)]. These show no equivalent effect, confirming this explanation.

Unless displacing strands are deliberately designed otherwise, it is likely that accidental hairpins will form, with the frequency increasing with strand length. The

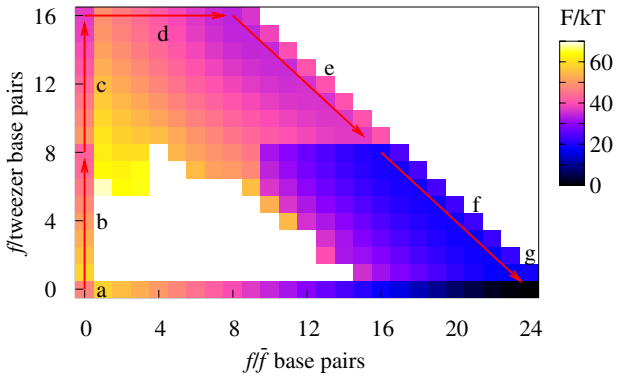

FIG. 4 (color online). Free energy F plotted as a function the number of $f/\bar{f}$ and f/tweezer base pairs for DNA tweezers at 300 K. White areas indicate regions of high free-energy relative to their environment that were unsampled.

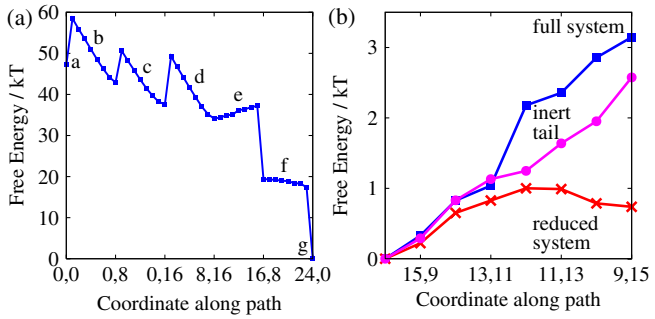

FIG. 5 (color online). (a) Free-energy profile along the one-dimensional pathway indicated in Fig. 4. Coordinates indicate the number of $f/\bar{f}$ and f/tweezer base pairs. (b) The displacement process "e" in more detail. Squares represent the original system, circles a system with the tail of $\bar{f}$ unable to form a hairpin, and crosses a system with the last eight bases of $\bar{f}$ and most of the f/β arm removed (see text).

nearest-neighbor model of SantaLucia [24] predicts that hairpins with stems of three base pairs and short loops are marginally stable at 300 K, supporting the suggestion that they can influence free-energy profiles. Furthermore, these hairpins will form either at the start or end of displacement, when long single-stranded regions are available. As a consequence, hairpin formation will generally constitute a free-energy barrier in the middle stages of displacement, thereby slowing down the process.

Second, steric effects contribute to the free-energy increase. On binding to the toehold of f, the unbound end of $\bar{f}$ has its conformational freedom restricted by the rest of the tweezers. As displacement begins, a second single-stranded region is formed, causing further steric restrictions. As more bases are displaced, the single-stranded regions are drawn into the body of the tweezers, causing additional steric restriction as illustrated in Fig. 3(e). Computer simulations of a reduced system in which the final eight bases of $\bar{f}$ (which are not involved in displacing α) and all but the first base pair of the f/β duplex were removed (details in Ref. [18]) show a significantly flatter landscape after the initial penalty for forming two single-stranded regions, confirming this explanation [Fig. 5(b)]. By contrast, the displacement of β by $\bar{f}$ reduces the amount of steric clashes as the tweezer unit is further separated from the f and $\bar{f}$ strands with each step, leading to a decrease in free energy during the displacement.

Many of the features of the free-energy landscape—the sharp initial rise upon forming the first base pairs, or even the more subtle effects of hairpin formation and excluded volume on the displacement steps—are sufficiently generic that they would survive the inclusion of additional chemical detail. Future model development will include the addition of base heterogeneity effects and the explicit effects of salt concentration, but even at the current level, we believe that our model will be particularly useful to study the design and operation of DNA nanomachines. Furthermore, we anticipate many potential applications for biologically relevant rearrangement transitions, such as the formation of cruciform DNA [25].

In summary, we have introduced a coarse-grained model of DNA which captures its thermodynamic and structural properties, representing stacking, duplex, and hairpin transitions consistently for the first time. The model makes possible the simulation of DNA nanostructure assembly and nanomachine operation, and has the potential to be extended into the biological domain.

[1] H. Yan *et al.*, Science **301**, 1882 (2003).
[2] J. Malo *et al.*, Angew. Chem., Int. Ed. **44**, 3057 (2005).
[3] N. C. Seeman, Nature (London) **421**, 427 (2003).
[4] R. P. Goodman *et al.*, Science **310**, 1661 (2005).
[5] P. W. K. Rothemund, Nature (London) **440**, 297 (2006).
[6] E. S. Andersen *et al.*, Nature (London) **459**, 73 (2009).
[7] S. M. Douglas *et al.*, Nature (London) **459**, 414 (2009).
[8] B. Yurke *et al.*, Nature (London) **406**, 605 (2000).
[9] S. J. Green *et al.*, Phys. Rev. Lett. **101**, 238101 (2008).
[10] T. Omabegho *et al.*, Science **324**, 67 (2009).
[11] T. E. Ouldridge *et al.*, J. Chem. Phys. **130**, 065101 (2009).
[12] M. Sales-Pardo *et al.*, Phys. Rev. E **71**, 051902 (2005).
[13] M. Kenward and K. D. Dorfman, J. Chem. Phys. **130**, 095101 (2009).
[14] F. W. Starr and F. Sciortino, J. Phys. Condens. Matter **18**, L347 (2006).
[15] K. Drukker, G. Wu, and G. C. Schatz, J. Chem. Phys. **114**, 579 (2001).
[16] E. J. Sambriski, V. Ortiz, and J. J. de Pablo, J. Phys. Condens. Matter **21**, 034105 (2009).
[17] E. J. Sambriski, D. C. Schwartz, and J. J. de Pablo, Biophys. J. **96**, 1675 (2009).
[18] See supplementary material at http://link.aps.org/supplemental/10.1103/PhysRevLett.104.178101 for model and simulation details.
[19] S. Whitelam and P. L. Geissler, J. Chem. Phys. **127**, 154101 (2007).
[20] G. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[21] W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
[22] J. Holbrook *et al.*, Biochemistry **38**, 8409 (1999).
[23] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Phys. Condens. Matter **22**, 104102 (2010).
[24] J. SantaLucia, Jr. and D. Hicks, Annu. Rev. Biophys. Biomol. Struct. **33**, 415 (2004).
[25] R. R. Sinden, *DNA Structure and Function* (Academic Press Inc., London, 1994).
[26] P. J. Hagerman, Annu. Rev. Biophys. Biophys. Chem. **17**, 265 (1988).
[27] M. C. Murphy *et al.*, Biophys. J. **86**, 2530 (2004).
[28] J. Bath and A. J. Turberfield, Nature Nanotech. **2**, 275 (2007).
[29] S. Kumar *et al.*, J. Comput. Chem. **13**, 1011 (1992).