

DNA nanotweezers studied with a coarse-grained model of DNA

Thomas E. Ouldridge¹, Ard A. Louis¹, and Jonathan P. K. Doye²

¹*Rudolf Peierls Centre for Theoretical Physics, 1 Keble Road, Oxford, UK OX1 3NP, UK*

²*Physical & Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford, OX1 3QZ, UK*

(Dated: November 3, 2009)

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.

PACS numbers: 87.14.gk,87.15.A-,81.07.Nb,34.20.Gj

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].

The free energy of hybridization can also be harnessed in artificial molecular machines. The simplest designs, such as DNA ‘tweezers’ [8], use alternating addition of two complementary 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 a DNA track, creating DNA walkers that undergo directional motion [9, 10].

Computer simulations of these DNA nanosystems would provide highly desirable insight into the details of the processes involved in assembly or mechanical cycles. Unfortunately, the system sizes and time scales involved make all-atom simulations prohibitively expensive. Instead, models that coarse-grain out the microscopic details must be used. In the remainder of this letter we introduce a coarse-grained model designed to reproduce generic DNA behaviour in both single- and double-stranded states, as well as the fundamental assembly transitions. We then demonstrate the use of the model for DNA nanostructures by simulating a full cycle for DNA tweezers.

A number of other coarse-grained models of DNA have been suggested in recent years. Non-helical 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

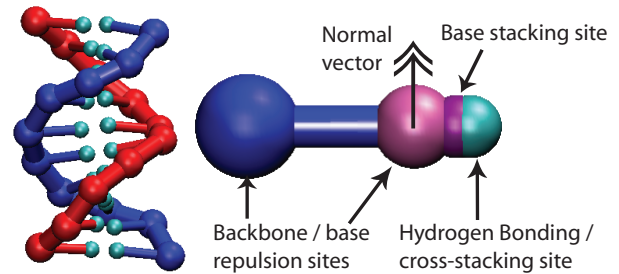


FIG. 1: (colour online) A duplex as represented by the model, and a detailed view of a nucleotide.

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 take a ‘top-down’ approach to DNA modeling, aiming to capture the generic properties of DNA that are important for assembly rather than reproducing all of the microscopic structural details. DNA is modelled as a string of rigid nucleotides, as depicted in Fig. 1, with one interaction site to represent the backbone and three for the base. The plane of the base is indicated by an additional ‘normal vector’. The backbone sites are connected via FENE springs, and act as soft repulsion centres (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 sites. Consecutive bases attract each other with a minimum at approximately 3.4 \AA , shorter than the equilibrium FENE spring length of approximately 6.5 \AA . We modulate this interaction according to the relative alignment of the normal vectors, and the alignment of the normal vectors with the inter-site vector. Thus the system is driven towards forming helical stacks of coplanar

bases: right handedness is imposed by setting the attraction to zero if the bases stack left-handedly.

Hydrogen bonding is represented by an attraction between hydrogen bonding sites of complementary bases, modulated by factors favouring co-linear 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. Firstly, 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, $[Na] = 500$ mM, where the Debye screening length is short (~ 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]. Due to the system’s simplicity and the efficiency of the algorithm, denaturation and hybridization of short duplexes can be observed without biasing the ensemble. To gain accurate statistics, however, we use umbrella sampling techniques [20] to characterize the basic DNA transitions.

The simplest of these is single-stranded stacking, in which 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 $\Delta H^{stack} = -5.6$ kcal mol $^{-1}$ and entropy of $\Delta S^{stack} = -16.6$ kcal mol $^{-1}$ K $^{-1}$, consistent with the experiments of Holbrook *et al.* [22].

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) obtained from the nearest neighbor model of SantaLucia [24], which is able to accurately predict experiments, for strands consisting of ‘average bases’ (defined by averaging over the parameters for all possible complementary base pair steps). Fig. 2 shows that our model is in excellent agreement with the predictions for T_m over a range of duplex lengths. Importantly, transition widths are also consistent 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-neighbour model by approximately 3 K (less than 1% of the absolute temperature), but importantly captures the dependence on loop (Fig. 2) and stem length (not shown).

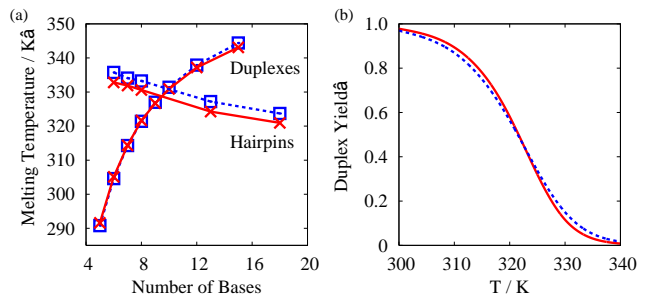


FIG. 2: (colour online) (a) Comparison of melting temperatures as computed for our model (crosses connected by a solid line) and predicted by the nearest-neighbour model [24] (squares connected by a dashed line) for duplexes as a function of the single-stranded length, and hairpins as a function of loop 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-neighbour model (dashed line).

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 160 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 18.2 Å (we define model length scales so that the average rise per base pair at 300 K is 3.3 Å). These values compare favourably 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 $\bar{f}\bar{f}$ 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. 3c). The additional 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 $\bar{f}\bar{f}$ 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

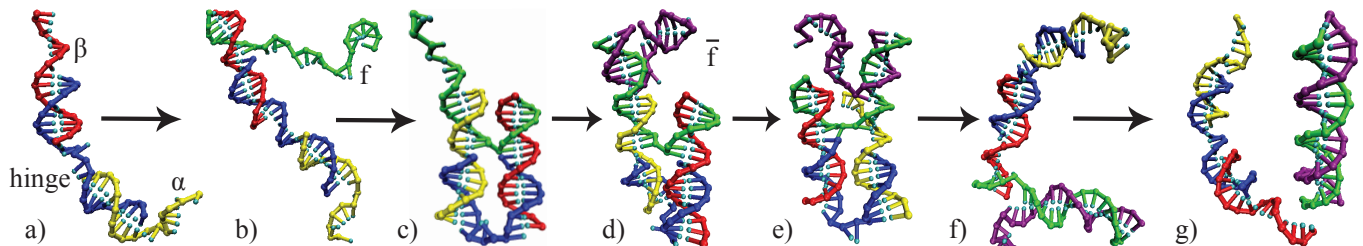


FIG. 3: (colour online) Simulation snapshots showing stages of operation of DNA tweezers. a) Tweezers initially open. b) Fuel (f) is added and binds to one arm (β). c) Fuel binds to the second arm (α) and closes the tweezers. d) Antifuel (\bar{f}) is added and binds to the toehold of the fuel. e) Antifuel 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.

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 $4.19 \times 10^5 \text{ nm}^3$ (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 $47.12 \pm 0.21 kT$ along this path; simulations of f and \bar{f} in isolation (displayed along $y = 0$ in Fig. 4) give $47.27 \pm 0.11 kT$, the agreement supporting the accuracy of our calculations.

The gross features of the free energy landscape are as expected. Duplex formation is highly cooperative; the pairing of two strands involves a high entropic cost for forming the first base pair, then a downhill slope in free energy as additional bonds are formed [17]. This is reflected in Fig. 5 by stages ‘b’, ‘c’ and ‘d’ which essentially

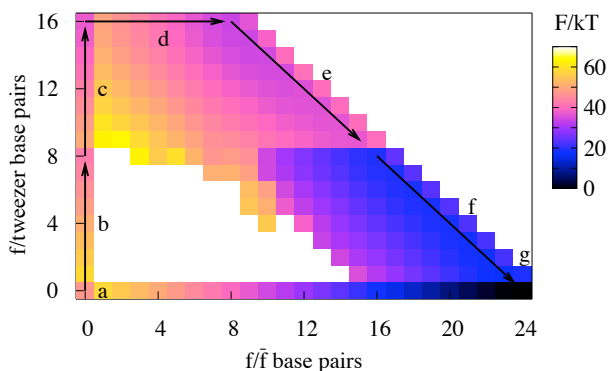


FIG. 4: (colour online) Free energy F plotted as a function of the number of f/\bar{f} and f /tweezer base pairs for DNA tweezers at 300 K. White areas indicate high free-energy regions that were unsampled.

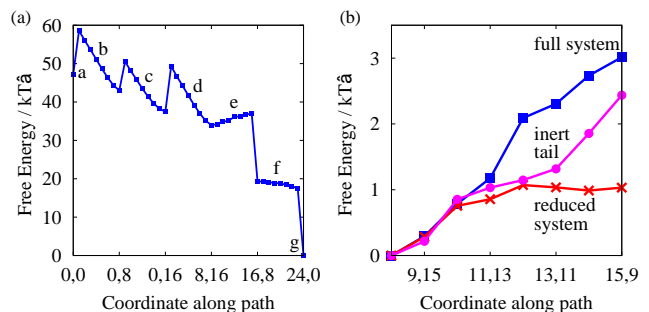


FIG. 5: (colour online) (a) Free energy profile along the one-dimensional pathway indicated in Fig. 4. Coordinates indicate the number of f/\bar{f} and f /tweezer base pairs. (b) The displacement process ‘e’ in more detail. Squares represent the original system, circles a system with the tail of f unable 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).

involve duplex formation. The large 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.

Computer simulations allow for a detailed inspection of processes like displacement. Thus, Fig. 5 shows that there is actually an increase in free energy of $\sim 3 kT$ 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 unconstrained 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 dis-

placement of the second.

Two effects help to explain the increase in free energy during the displacement of α . Firstly, \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 small hairpins will form, with the probability of accidental hairpins increasing with the length of the strand. The nearest-neighbour 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 of our simulations 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.

The second reason for the increase of free energy comes from steric effects. On binding to the toehold of f , the unbound end of \bar{f} has its conformational freedom restricted by the presence of 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 re-

duced 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 even if much more chemical detail was included in the simulations. 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 new coarse-grained model of DNA which reproduces its thermodynamic and structural properties, representing single-stranded 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* **421**, 427 (2003).
 [4] R. P. Goodman *et al.*, *Science* **310**, 1661 (2005).
 [5] P. W. K. Rothmund, *Nature* **440**, 297 (2006).
 [6] E. S. Andersen *et al.*, *Nature* **459**, 73 (2009).
 [7] S. M. Douglas *et al.*, *Nature* **459**, 414 (2009).
 [8] B. Yurke *et al.*, *Nature* **406**, 605 (2000).
 [9] S. J. Green, J. Bath, and A. J. Turberfield, *Phys. Rev. Lett.* **101**, 238101 (2008).
 [10] T. Omabegho, R. Sha, and N. C. Seeman, *Science* **324**, 67 (2009).
 [11] T. E. Ouldridge, I. G. Johnston, A. A. Louis, and J. P. K. Doye, *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** (2009).
 [17] E. J. Sambriski, D. C. Schwartz, and J. J. de Pablo, *Bio-phys. J.* **96**, 1675 (2009).
 [18] See EPAPs Document No. [?] 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. Comp. Phys.* **23**, 187 (1977).
 [21] W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, 1984).
 [22] J. Holbrook, M. Capp, R. Saecker, and M. Record, *Biochemistry* **38**, 8409 (1999).
 [23] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, arXiv:0910.1201.
 [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., 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, *Nat. Nanotechnol.* **2**, 275 (2007).
 [29] S. Kumar *et al.*, *J. Comput. Chem.* **13**, 1011 (1992).