

Coarse-graining DNA for simulations of DNA nanotechnology

Jonathan P. K. Doye,^{*a} Thomas E. Ouldridge,^b Ard A. Louis,^b Flavio Romano,^a Petr Šulc,^b Christian Matek,^b Benedict E. K. Snodin,^a Lorenzo Rovigatti,^c John S. Schreck,^a Ryan M. Harrison^a and William P. J. Smith^a

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXX 20XX

First published on the web Xth XXXXXXXXXX 200X

DOI: 10.1039/b000000x

To simulate long time and length scale processes involving DNA it is necessary to use a coarse-grained description. Here we provide an overview of different approaches to such coarse graining, focussing on those at the nucleotide level that allow the self-assembly processes associated with DNA nanotechnology to be studied. OxDNA, our recently-developed coarse-grained DNA model, is particularly suited to this task, and has opened up this field to systematic study by simulations. We illustrate some of the range of DNA nanotechnology systems to which the model is being applied, as well as the insights it can provide into fundamental biophysical properties of DNA.

1 Introduction

DNA is one of the fundamental molecules of biology, enjoying iconic status in its role as the information storage medium for the genome. More recently, DNA has also become a startlingly successful nanoengineering material, with the exponents of DNA nanotechnology having developed an impressive array of DNA nanostructures¹ and nanodevices.² For both these uses of DNA, its physico-chemical properties are vital to its role.

For example, the intra- and intermolecular interactions that drive base pairing and base stacking are responsible for DNA's double helical structure. Furthermore, the selectivity of Watson-Crick base-pairing and the efficiency with which double helices can self-assemble underlie both fundamental genomic processes, such as replication and transcription, and the programmability of DNA that has been exploited so successfully by DNA nanotechnology. On longer length scales, the mechanical properties of the double helix are important both to how DNA can be manipulated in the cellular environment, such as through the packaging of DNA by complexation with proteins and the control of its global topology by enzymes that modify its local twist, and in imparting rigidity to DNA nanostructures.

Experimentally, there is an increasing wealth of information on DNA's biophysical properties, particularly more recently from single-molecule experiments involving active manipula-

tion by optical and magnetic tweezers^{3,4} or by passive fluorescent probing of single molecules as they undergo dynamic processes.^{5,6} However, the detailed microscopic mechanisms underlying the observed behaviour in these experiments are not always clear.

DNA nanotechnology is also a relatively young field. It originates from the 1980s when Nadrian Seeman first proposed making artificial objects out of DNA that could form by self-assembly.⁷ In the last decade this field has seen very rapid growth, both in terms of the size of the literature, and the complexity of the objects that have been produced. The structures typically consist of double-helices joined together by single-stranded sections or by junctions at which strands cross over between helices, but the technology has now developed to such an extent that objects of seemingly any shape can be assembled.¹ One of the distinctives of these DNA nanostructures is that they can be addressed with nanometer precision, and so one recent theme has been the decoration of such structures with, for example, nanoparticles⁸ and other biomolecules⁹ in order to create functionality or to use as substrates for single-molecule experiments. A second attractive feature is that structures can be made that are able to respond, usually to the presence of other nucleic acid strands but also to other molecules,¹⁰ to pH¹¹ and to light.¹² Such devices have included nanotweezers,¹³ openable boxes,¹⁴ walkers that can move along tracks¹⁵ and responsive drug delivery capsules.¹⁰

The most basic design principle for DNA nanostructures is simple; namely, that the target structure should be the global free-energy minimum, and this can usually be achieved if the target has the most base pairs. However, how such structures, particularly the larger and more complex ones, which can be made up of tens of thousands of nucleotides and hundreds of

^aPhysical & Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford, OX1 3QZ, UK.

^bRudolf Peierls Centre for Theoretical Physics, 1 Keble Road, Oxford, OX1 3NP, UK.

^cDipartimento di Fisica, Università di Roma La Sapienza, Piazzale A. Moro 2, 00185 Roma, Italy.

different strands, are able to self-assemble is still somewhat of a mystery. Moreover, there are relatively few guiding principles for how to prevent the systems from getting stuck in kinetic traps, and hence to improve yields. For DNA nanodevices, well-established nearest-neighbour thermodynamic models of duplex and hairpin stability¹⁶ can help provide insight into their action. However, the rates of the different processes involved are also important. Furthermore, these devices can involve non-trivial multi-strand complexes with non-trivial internal loops or pseudoknots that can be subject to internal or externally applied stresses and whose stabilities are not yet captured by these thermodynamic models.

The above issues suggest that there could be a significant potential role for molecular simulations in DNA nanotechnology: firstly, in providing an improved microscopic understanding of basic DNA biophysics relevant to DNA nanotechnology; secondly, in visualizing the mechanisms of self-assembly for DNA nanostructures and of the action of DNA nanodevices, and thirdly in aiding the rational design and optimization of DNA nanotechnology systems using the potential quantitative insights into their thermodynamics and dynamics. So far, however, the contribution of molecular simulations to DNA nanotechnology has been relatively minor.

One significant obstacle is that all-atom simulations are not yet capable of probing the large system sizes and long time scales that are typically relevant to DNA nanotechnology systems. Therefore, to progress the use of simulations in this field, there is a need for “coarse-grained” models that provide a simpler representation of DNA, but one which hopefully retains enough of the essential physics to allow realistic modelling of DNA systems on long time and length scales. However, until relatively recently, there were few such coarse-grained models, and coarse-graining has been more explored for other biomolecules¹⁷ such as proteins¹⁸ and lipids.¹⁹ This is perhaps surprising given that, in some ways, DNA is an easier target for coarse-graining, certainly when compared to proteins, because of the relatively limited number of ways the bases interact — the physical behaviour of DNA is dominated by the stacking of the bases and the pairing of complementary bases through hydrogen-bonding. Only in the last few years have the number of coarse-grained DNA models increased significantly,^{20–22} but only a few of these have the potential to realistically model self-assembly processes associated with DNA nanotechnology systems. Notably, the first published example of such an application was only in 2010 when nanotweezers similar to those developed by Yurke *et al.*¹³ were simulated.²³

Here we provide a selective overview of coarse-grained models of DNA in Section 2 focussing down on those that have the requisite properties to allow the simulation of DNA nanotechnology and highlighting the oxDNA model that we have recently developed for this purpose in Section 3. In Sec-

tion 4 we discuss some of the technical issues that are relevant when computing the thermodynamic and dynamic properties of such models by molecular simulation, before illustrating the power and range of applicability of oxDNA by showing a number of nanotechnological and biophysical systems to which the model has been applied in Section 5.

2 Coarse-grained DNA models: a selective overview

When seeking to understand a particular property of DNA by theory or simulation, one of course wants to choose a level of description that captures enough of the physics to allow the question of interest to be addressed, whilst also remaining sufficiently simple that such a calculation is tractable. For this reason, there is a whole spectrum of approaches to describe DNA involving differing levels of detail.

At one end of this spectrum are polymer theories such as the worm-like chain model, which describes DNA as a polymer with a certain bending modulus (and sometimes finite twist and extensional moduli). This model is particularly suited to looking at the elastic properties of DNA on long length scales. For example, it provides a good description of the response of double-stranded DNA to low forces, capturing its entropic elasticity.²⁴ However, such a description begins to break down for mechanical properties that involve more extreme perturbations in the structure. For example, the worm-like chain model can describe the “*j*-factors” for the cyclization of long DNA duplexes, but fails for short duplexes where cyclization is most likely achieved by local buckling rather than homogeneous bending.²⁵

At the other end of the spectrum, certain properties can probably only be understood using models that have an all-atom representation of DNA, because the property reflects the detailed local geometry of the molecule. Although the achievements of all-atom simulations of DNA are becoming increasingly impressive,^{26,27} due both to advances in the force-fields used and to greater computer power, such methods are inevitably limited to either relatively short time scales or relatively small system sizes.

In between these two extremes are a whole range of different coarse-grained models of varying complexity. As we wish to consider models that would allow DNA nanotechnology systems to be studied, in particular the self-assembly, structure and mechanical properties of DNA nanostructures, and the action of DNA nanodevices, we should first consider the requirements this imposes on such a model. The first and most basic requirement is that the model has a realistic description of the three-dimensional geometry of DNA. Secondly, the self-assembly of virtually all such DNA nanostructures and the action of most of the DNA nanodevices is driven

by hybridization, and so it is important that this transition is well described. On a thermodynamic level this includes being able to reproduce the melting points and transition widths of duplex DNA. Thirdly, a good description of the mechanical properties of both single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) is required if the structural and mechanical properties of DNA nanostructures are to be reproduced, in particular, the rigidity of the duplex and the relative flexibility of ssDNA. The latter is particularly important, as virtually all DNA nanostructures make use of the ability of the DNA backbone to bend back on itself at junctions connecting different double helical sections of the structures. It can also be important to have a good description of the extensibility of ssDNA, since sometimes the operation of nanodevices involves externally applied or internally generated tension.^{28,29} Fourthly, it must be feasible to simulate the long time scales associated with diffusion of strands in solution and with significant conformational rearrangements.

These considerations rule out many possible models. For example, all-atom models are simply too computationally expensive to allow the time scales associated with self-assembly to be simulated. The same is probably true of those coarse-grained models that retain an explicit description of the solvent.³⁰ Similarly, the need to describe hybridization rules out those coarse-grained models which are intended just to describe dsDNA and do not dissociate.^{31–37} Therefore, the minimal unit must at most be at the level of a single nucleotide. It is interesting to note that Lankaš *et al* found that the level of the base also provides a more appropriate level to capture the local dynamics of DNA than that of the base pair.³⁸

Models that have too reduced a representation of DNA geometry will also be unsuitable. For example, the Peyrard-Bishop family of models have been extensively used to study melting processes in DNA, their simplicity allowing particularly long length scales to be explored,³⁹ but their lack of three-dimensional structure makes them inappropriate for our task. Similarly, although “ladder models” in which dsDNA is not helical,^{40–44} might be able to give qualitative insights into the nature of self-assembly for DNA-like polymers that have complementary base-pairing interactions,^{45,46} they are not suitable for detailed quantitative analysis. For example, the mechanical properties of dsDNA in the model are too anisotropic; in particular, fluctuations out of the plane of the “ladder” are too facile. However, their simplicity allows particular large systems, such as DNA-coated nanoparticles, to be studied.⁴⁷

Thus, these exclusions leave coarse-grained DNA models that can undergo melting and have a reasonable description of basic DNA geometry. Although the development and use of coarse-grained models of DNA has generally lagged behind that for other biomolecules, such as proteins or lipids, there are a rapidly increasing number of models at this level.^{23,48–70} The

distinctions between these models that we wish to highlight are the “philosophies” behind the models and the ways that the component interactions are represented.

The first major question is to which data should the model be fitted, and there are two main approaches. The first, which we term the “bottom-up” approach, is to fit the model to results (typically correlation functions) from a finer-grained model, which is usually an empirical all-atom model (e.g. AMBER, CHARMM),^{35–37,63,64,67–69} although fitting to energies obtained from electronic structure calculations (albeit in the absence of water)⁶⁵ has also been performed. The second “top-down” approach is to directly fit the parameters of a physically motivated potential to reproduce measured experimental properties, e.g. melting transitions, persistence lengths, and elastic moduli.

Although the bottom-up approach may provide a more direct connection to models with finer detail, there are a number of potential issues to bear in mind with the bottom-up approach. Firstly, it is dependent on the quality of the fine-grained description, but for many properties in which one might be interested, e.g. DNA melting points, it is often unknown how well the fine-grained description reproduces these properties, simply because they are currently too computationally expensive to calculate at that level. Secondly, although some methods of coarse-graining may have more rigour than others, there is no unique approach to achieve the coarse-graining, and “representability” problems are a ubiquitous feature of such coarse-graining.⁷¹

The second major question is how to represent the interactions. It is probably most common to use forms that aim to capture some of the basic structural features of the interactions. Therefore, often the potential is broken up into terms that represent base stacking, the hydrogen-bonding between complementary base pairs, a backbone potential, excluded volume, electrostatic interactions between charged groups, and perhaps cross-stacking (interactions between diagonally opposite bases in the duplex), coaxial stacking (stacking interactions between bases that are adjacent in a helix, but are not neighbours along a chain) and a solvation term.

There are a number of things to be born in mind when considering the interactions in a coarse-grained model. Firstly, the interaction terms are in principle free energies, not just energies.^{71,72} By definition, coarse-graining, by reducing the number of degrees of freedom, reduces the total entropy of the system. However, this is not a problem in itself, because we are usually concerned with entropy differences, e.g. between stacked and unstacked bases, rather than absolute entropies. Nevertheless, there is no guarantee that these relative entropy differences will be preserved by the coarse-graining, and so consideration should be given to the potential entropic components, and hence temperature dependence, of the effective interactions. Most coarse-grained potentials ignore this issue

and are temperature independent.

For some degrees of freedom that are coarse-grained away, there is likely to be little effect on the relative entropies. For example, the internal vibrations of the bases are likely to be very similar in the stacked and unstacked states, and so the loss of these internal degrees of freedom is likely to be of little thermodynamic consequence.

Upon stacking, the bases lose entropy due to the more restricted nature of both the relative positions of the centres of mass and the relative orientations of the bases. If a model does not contain all these potential geometric sources of entropy, for example, if, as many models do, a base is represented by a single site with no orientational degrees of freedom,^{48,49,51,53–55,59–62,65} or if these degrees of freedom are not captured correctly, then it is likely that an entropic term in the stacking interaction would be needed to compensate.

Furthermore, most models also coarse-grain away the water, and the effects of this on the thermodynamics is much harder to predict. Stacking is partly driven by the hydrophobic character of the bases, but correctly describing hydrophobicity is far from straightforward and is known to often have a strong entropic component.⁷³ Therefore, it would be unsurprising if an entropic component to the stacking would be required to account for the contribution from hydrophobicity.

The second issue that needs careful consideration when designing a coarse-grained model is the detailed forms used to describe the interactions, with physically sensible choices of these forms being a prerequisite for a good model. Here, we highlight what we think are key aspects of any model.

One important feature is the origin of the helicity of dsDNA and the flexibility of ssDNA. Of course, all models should have the double helical B-DNA geometry as the global free-energy minimum of complementary strands below their melting point. One way to achieve this is to have the B-DNA geometry as the minimum of all the individual interactions. However, this choice may have significant consequences for the structure of ssDNA. In particular, this approach can lead to a backbone torsional potential that imposes helicity on a single DNA strand, and hence to an unphysically rigid helical geometry for ssDNA.^{51,53,54} In fact, DNA's helicity results from the combined constraints of the backbone and the stacking interaction, and in particular, the difference in distance between the separation of stacked bases (3.4 Å) and that typical for the separation of bases along the DNA backbone (approximately ~6.5 Å), with a (double) helix being the most favourable way of satisfying both.⁷⁴ Thus, a model that achieves duplex helicity through these two distance scales has the advantage that ssDNA when unstacked is flexible and can kink sharply, although it can still also adopt a helical geometry when stacked.

Interestingly, a recent coarse-grained model for dsDNA that was derived by rigorous fitting to extensive all-atom simulations indicated that the individual interactions are not all op-

timally satisfied in the B-DNA structure.³⁷ The “frustration” that results may be key to explaining non-local effects of sequence on the double helical structure,³⁷ and may also be important to capture the thermodynamic destabilization of the duplex that arises from mismatches.⁶⁶

A second important feature is to capture the anisotropic nature of the interactions between bases, both for stacking and hydrogen-bonding. As noted already, if a base is represented by a single site without orientational degrees of freedom, there is the potential problem that it will be hard to capture the entropy loss associated with stacking. Furthermore, Watson-Crick base-pairing interactions should be able to occur to one other base, when the bases are coplanar and the chains are anti-parallel. However, if the description of the hydrogen-bond interaction between base sites is isotropic,^{51,53–55} there is the possibility that the “single-valent” nature of Watson-Crick base pairing can be broken. For example, this is exactly what Florescu and Joyeux found to occur for the 3SPN model of de Pablo and coworkers^{51,53,54} for a polydA-polydT duplex. The two strands spontaneously slip by half a base-pair rise with respect to each other so that each base can bind two bases on the complementary strand.⁷⁵ Note that this effect is normally prevented by the heterogeneity of a sequence.

In summary, although there are now a considerable number of coarse-grained DNA models available, most are less-suited to studying the self-assembly processes associated with DNA nanotechnology, because either they do not meet all the requirements set out at the beginning of this section or their behaviour has not yet been sufficiently well characterized. Of course, this is not to say that these models do not have their own domains of applicability where they can be productively used. OxDNA, the coarse-grained model created in our groups in Oxford, was developed with such nanotechnological applications explicitly in mind, and it is our contention that it is currently the most well-suited model for simulating DNA nanotechnology; it is also the model that has been applied to the most DNA nanotechnology systems.^{23,28,29} Therefore, for the rest of the article we focus on this model.

Nevertheless, we note that if one is only interested in the structural properties of DNA nanotechnology systems, other simulation approaches can potentially be used. For example, if the system is sufficiently small it may be possible to use all-atom simulations, as has been done for a few examples.^{76,77} Also for structures in which all the bases are hybridized, coarse-grained models that describe the double-helical structure well could potentially be applied, as long as the effects of junctions can be incorporated into the model. For example, the rigid base-pair model of Ref. 31 has been adapted to probe the structure of DNA origamis,⁷⁸ and the CanDo package, which assumes that double helices behave like elastic rods, is a useful resource for predicting DNA origami structure and flexibility.^{79,80}

3 The oxDNA model

In the oxDNA model each nucleotide is represented by a rigid nucleotide that consists of a set of collinear interaction sites and a vector that is perpendicular to the notional plane of the base (see Fig. 1(a)). The aim of the vector is to capture the planarity of the base through the orientational dependence of the interactions rather than through additional sites. This orientational dependence allows the model to represent the coplanar base stacking and both the linearity of hydrogen-bonding and the edge-to-edge character of the Watson-Crick base pairing. Furthermore, it is also used to capture more subtle features of the interactions that are consequences of more detailed structural features of DNA that are not present in the model, for example, the right-handed character of the double helix and the anti-parallel nature of the strands in the helix.

Note also that, in contrast to a number of models, there is a single backbone site, rather than separate sites representing the sugar and phosphate groups. We feel that this is a reasonable approximation for most of the systems in which we are interested, but it may affect detailed structural properties for systems where the backbones come in close proximity. One practical advantage of this choice is that it allows us to maintain the pairwise nature of the internucleotide interactions without the need for three-body terms to enforce the geometry of the backbone. Furthermore, coarse-graining the whole nucleotide into a single rigid body allows us to ignore the internal motions of the nucleotides, thus significantly speeding up sampling.

The form of the internucleotide potential is

$$V_{\text{oxDNA}} = \sum_{\langle ij \rangle} (V_{\text{b.b.}} + V_{\text{stack}} + V'_{\text{exc}}) + \sum_{i,j \notin \langle ij \rangle} (V_{\text{HB}} + V_{\text{cr.st.}} + V_{\text{exc}} + V_{\text{cx.st.}}), \quad (1)$$

where the first sum is taken over all nucleotides that are adjacent along the backbone of a strand and the second sum is taken over all other pairs of nucleotides. The different terms represent backbone connectivity ($V_{\text{b.b.}}$), excluded volume (V_{exc} and V'_{exc}), hydrogen bonding between complementary bases (V_{HB}), stacking between adjacent bases on a strand (V_{stack}), cross-stacking ($V_{\text{cr.st.}}$) and coaxial stacking ($V_{\text{cx.st.}}$). The nucleotides that participate in these interactions within the double-helical state are schematically shown in Figure 1(b). The excluded volume and backbone interactions are isotropic, whereas all other interactions depend on the relative orientations of the nucleotides as well as the distance between the interaction sites. The detailed forms for all of the terms in Eq. 1 are available in Ref. 81. Note that these forms are not attempting to provide a chemically realistic description of the interactions, but instead a physically realistic description of effects of the interactions.

We applied a top-down approach to parameterize the

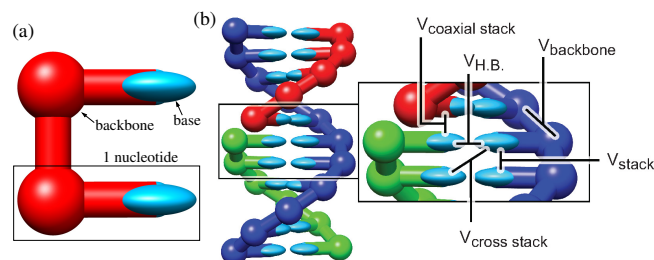


Fig. 1 (a) A representation of the rigid nucleotides that are the basic unit of the oxDNA coarse-grained model. The bases are represented by an ellipsoid to reflect the orientational dependence of the interactions. (b) Three strands in an 11-base-pair double helix with stabilising interactions indicated in the inset. All nucleotides also interact through short-ranged excluded volume repulsions.

oxDNA model. Loosely speaking, the structure is determined by the positions of the minima in the interaction potentials, the thermodynamics by the well depths and force constants, and the mechanical properties also by the force constants, and all three types of data have been used to fit the parameters of the model. In particular, for the thermodynamics, we have fitted the model to reproduce the transition temperatures and widths for the melting of short duplexes as predicted by the SantaLucia model,¹⁶ and results for the stacking transition of ssDNA.⁸² Similarly, for the mechanics we have fitted to the persistence lengths of ssDNA and dsDNA and the elastic moduli of dsDNA. We note that for the fitting of the model it would have been very useful to have more unequivocal experimental data on the physical properties of single-stranded DNA, particularly for the stacking transition.

As noted in the previous section, the effective interactions in a coarse-grained model should in principle be free energies. However, if the coarse-graining maintains the correct relative entropies between the states of interest, the entropic components of the interactions would be zero. As we have tried to retain all the geometric sources of entropy relevant to hybridization (e.g. the orientational degrees of freedom of the bases and the flexibility of the single strand), we attempted to keep the interactions temperature-independent as far as possible. In the end, it only proved necessary to introduce temperature dependence into the stacking interactions (and only to a relatively small degree) in order to obtain a good fit to the thermodynamic data. This suggests that our model has been relatively successful in retaining the relevant geometric degrees of freedom; of course, we are unable to retain the entropic components of the interactions related to the solvent, and so the need to introduce temperature-dependence into oxDNA's stacking interactions, which have a significant hydrophobic component, might reflect this source of entropy.

There are currently two parameterizations of the oxDNA model available. Firstly, there is an “average-base” pa-

parameterization in which the interactions for the different bases are identical except for the hydrogen-bonding term, for which bonding only occurs between Watson-Crick base pairs.^{23,66} In the second parameterization, the interaction strengths for stacking and hydrogen-bonding have been allowed to vary in order to capture the sequence-dependence of DNA thermodynamics.⁸³ However, the development of this sequence-dependent version has not made the average-base model redundant, since it is often useful to examine the generic behaviour of DNA, unobscured by sequence-dependent effects. Indeed, most of the published applications of oxDNA have been reported for the average-base parameterization.^{23,28,29,84–89} By contrast, the sequence-dependent parameterization is particularly useful when we want to compare in detail to a particular experimental system.

As with any coarse-grained model, it is important to be open about oxDNA's limitations. Firstly, the model is parameterized for a specific salt concentration, namely 500 mM, and the electrostatics interactions between charged groups are not explicitly represented but instead their effects are incorporated into the excluded-volume interactions. This approach is reasonable because of the short-range nature of the Debye screening length at this salt concentration. Furthermore, these solution conditions are not untypical of the relatively high ionic strengths used in DNA nanotechnology experiments. In principle oxDNA could be reparameterised at different salt concentrations, but at lower salt it is not clear the short-ranged functional forms we employ would be sufficiently flexible. Some progress could perhaps be made by through a Debye-Hückel like description of the electrostatics, as has been done for a number of other DNA models.^{51,53,54,56,65,67} However, it should be kept in mind that further complications such as non-linear electrostatics also become increasingly important at low salt, see e.g. Refs. 90 and 91 for further discussion.

Secondly, even for the second parameterization of the oxDNA potential, the sequence-dependence is limited. All properties of the bases are assumed to be the same except for the strengths of the attractive interactions; for example, all bases are still of the same size. For this reason, the model is unlikely to be able to address the sequence dependence of the detailed structure and elasticity of dsDNA.

Thirdly, the double helix in oxDNA is symmetrical with equal sizes for the major and minor grooves (Fig. 1(b)). For many applications this deficiency will be of little importance, but it will affect systems where the detailed geometry of the backbone is significant. For example, major and minor grooving is needed to reproduce the correct relative angles between crossovers in DNA origami when both staple and scaffold strand crossovers are present. For this reason, we are in the process of parameterizing a version of the model that includes the correct grooving.

We note that simulation codes incorporating the oxDNA po-

tentials and all the algorithms presented in the following section are available from the oxDNA web-site.⁹²

4 Simulating coarse-grained models

DNA molecules in solution undergo diffusive motion. In molecular dynamics simulations of all-atom models, this diffusional dynamics is achieved through collisions between the DNA and the solvating water molecules, which also act as a thermal and pressure reservoir. However, in virtually all coarse-grained DNA models, the solvent is modelled implicitly and so applying standard molecular dynamics to such models would generate artefactual ballistic dynamics for the DNA. Further, simulations would be isoenergetic, whereas sampling from the canonical ensemble is appropriate when models of dilute systems with implicit solvent are to be compared directly to experiment.⁹³

A solution to this problem is to use Langevin dynamics,⁹⁴ in which additional drag and random forces are added to Newton's equations of motion in order to achieve diffusional dynamics. The size and form of these forces can be chosen self-consistently so that the system samples from the canonical ensemble. In the limit of strong noise and frictional forces, inertia can be neglected in a formalism known as Brownian dynamics. A simpler approach is to use Andersen-like thermostats,⁹⁵ in which systems are evolved using Newton's equations between occasional 'collision steps' when particle momenta are resampled from the thermal distribution. On scales longer than the typical time for resampling momenta, motion of particles is diffusive (although it is ballistic on shorter times).

Langevin, Brownian and Andersen-like thermostats all produce diffusive motion and sample from the canonical ensemble. In most implementations, however, they neglect long-range hydrodynamic forces which (amongst other effects) accelerate the diffusion of large clusters of particles. As a consequence, relative diffusion rates of different size clusters will not scale correctly when systems are simulated using these approaches. It is possible to incorporate long-range hydrodynamic effects in Langevin or Brownian algorithms,⁹⁴ but generally this is not done due to the extra computational cost.

Issues with cooperative hydrodynamics highlight the problem of comparing to experimental time scales. Designing a model with mass, energy and length scales then automatically defines a unit of time, which in principle could be used to compare with experiment. However, many experimental measurements reflect processes which require overall diffusion of strands, and as we have pointed out most dynamical implementations do not reproduce the physical scaling of diffusion coefficients with system size. An added complication is that coarse-grained models tend to smooth out the microscopic roughness of energy landscapes, thereby accelerating

motion.⁹⁶ Given these caveats, it seems sensible to focus primarily on relative rates of similar processes when simulating coarse-grained models, where uncertainties in mapping time scales will tend to cancel out. We note that the point of coarse-graining is to make simulations faster, and so one can even deliberately accelerate dynamics by using artificially high diffusion coefficients. For oxDNA at least, we have shown that, as long as one is careful to compare relative rates, such an approach has no qualitative consequences for the process of duplex formation.⁸⁸ More generally, we expect this approach to be reasonable if there is a separation of the time scales between diffusion and the relevant internal rearrangements of the strands. Indeed, an alternative way to map time scales onto experiment is to simply scale the simulation time to match the experimental diffusion constant of a strand, although this approach becomes more complicated when the simulations involve strands of different length because coarse-grained simulations do not generally reproduce the physical scaling of the diffusion coefficient with size.

Monte Carlo is essentially a method to perform equilibrium sampling. However, if the aim is simply to reproduce diffusive motion and sample from the canonical ensemble, Monte Carlo methods using local moves can provide an alternative approach to simulate dynamics.^{97–100} However, if one were to apply a Monte Carlo algorithm that used “single particle” moves to oxDNA, the simulations would be quite inefficient, because a significant fraction of the moves that involve strongly interacting nucleotides would be rejected. As a result, the diffusion of strands would be particularly slow, and even more so if the strands were part of duplexes or higher-order aggregates. Thus, single-particle move Monte Carlo will drastically fail to reproduce the relative diffusion coefficients of different-sized aggregates.

One way to overcome this problem is to use a Monte Carlo algorithm that attempts moves of clusters of nucleotides. One such algorithm that we have found to be particularly useful is the Virtual Move Monte Carlo (VMMC) algorithm of Whitlam and Geissler¹⁰¹ (note that we actually use the variant in the appendix of Ref. 102). One of the advantages of this algorithm is that the clusters that are constructed reflect both the current configuration and the proposed move. As a result, the algorithm allows more efficient sampling for our DNA model than Monte Carlo with single-nucleotide moves. We note here that cluster algorithms work best when the system is dominated by pairwise potentials, as multi-body interactions cannot generally be used in the cluster growth procedure.

In principle, one can modulate the VMMC algorithm to try and capture relative diffusion rates of clusters of different sizes.¹⁰¹ Reproducing physical diffusion rates is far from simple, however, and may have consequences for efficiency as the techniques proposed involve rejecting moves that would otherwise be accepted. It is also difficult to directly compare simu-

lation times for systems of different sizes, or when an external force is applied in one case and not in another. Nonetheless, we have found that VMMC trajectories often give qualitatively similar results to Langevin and Andersen-like thermostats for oxDNA, suggesting that the kinetic results are not artefacts of a particular choice of algorithm. One possible approach to generate dynamical Monte Carlo trajectories would be to combine VMMC with simpler cluster algorithms, such as that of Bhattacharyay and Troisi.¹⁰³ VMMC could be used to capture internal relaxation of clusters and the simpler algorithm to drive overall diffusion.

In our simulations of oxDNA, we have used a variety of the above techniques. VMMC has proved to be extremely computationally efficient for simulating small systems (smaller than about 200 nucleotides) in which processes involving hybridisation or melting of duplexes are of interest. It is also easy to couple to rare-event methods such as umbrella sampling (see section 4.1), making it ideal for measuring the thermodynamics of DNA reactions. The efficiency of VMMC stems from the ability to make relatively large moves without losing precision, unlike methods that involve integrating equations of motion. We note that very large moves that could cause strands to pass through each other must be excluded, however, if topology preservation is required.

In general, we have used Langevin¹⁰⁴ and Andersen-like¹⁰⁵ algorithms (specifically designed for rigid bodies) to study reaction kinetics, as the dynamics is easier to interpret than for VMMC (even if more computational time is required to observe processes). For equilibrating structural properties of large systems such as DNA origamis or long duplexes, we have used the Andersen-like algorithm as it is most computationally efficient (VMMC is less effective at performing internal mechanical relaxation for large clusters, and generates many unhelpful moves of the entire system). Additionally, Langevin and Andersen-like thermostats are easier to parallelize than cluster-building Monte Carlo approaches, which are naturally more serial in character. Implementing the Andersen-like algorithm on graphical processing units has allowed us to treat extremely large systems of tens of thousands of nucleotides, and so sample the structural equilibrium of DNA origamis and reactions involving many strands.

4.1 Rare-event methods

Although coarse-grained models allow much longer time scales to be simulated than for all-atom models, many of the processes involving DNA in which one might be interested involve significant free-energy barriers, and so may still be hard to sample. For example, free-energy barriers can arise due to the loss of translational entropy when two species associate, or due to the necessity of breaking base pairs, or if a process is geometrically or topologically constrained.⁸⁴ In character-

ising such processes, rare-event simulation techniques are extremely useful; we shall first consider techniques designed to measure the equilibrium properties of such systems, before dealing with approaches for measuring kinetics.

4.1.1 Accelerating equilibration. Two methods of particular note for improving equilibration are umbrella sampling^{95,106} and parallel tempering.¹⁰⁷ Umbrella sampling enhances the equilibration of thermodynamic properties by artificially biasing the formation of intermediate configurations between two or more local free-energy minima, thereby increasing the rate of transitions between these minima. The sampling bias can then be corrected to yield an unbiased free-energy landscape for the process in question. Parallel tempering involves running simultaneous replica simulations at a range of temperatures, and occasionally swapping configurations between replicas with a probability that preserves the canonical distribution at each temperature. This approach can accelerate equilibration if reactions are faster at temperatures above or below the one in question. In general, we have found the precise control offered by umbrella sampling to be most useful in measuring thermodynamic properties of our systems, with parallel tempering an alternative when it is difficult to define an effective order parameter.⁸⁶

When applying such rare-event methods to the formation of a target structure from more than one species, it is usually most computationally convenient to consider the formation of a single target, e.g. one DNA duplex from two complementary strands. However, the statistical distribution of various clusters of strands within a simulation volume obtained from a single-target simulation is different from the distribution that would be obtained from a bulk simulation of the same model (at the same concentration and temperature).^{93,108} Physically, the difference arises because concentration fluctuations are suppressed. For example, in the case of a single duplex, there would always be one strand of each type in the simulated volume, whereas for the same volume in a bulk system there are many other possibilities. Assuming that the concentration of duplexes in the simulated system is representative of bulk leads to a melting transition that is much narrower (as temperature is varied) than would be observed in bulk (the width is approximately 50% larger in bulk). For non-self complementary duplexes, the duplex concentration is also systematically overestimated at all temperatures, leading to an error in the melting temperature of several Kelvin, depending on the width of the transition. To avoid this problem, one could of course choose to simulate a larger system that is able to form multiple copies of the target structure. However, the finite-size effects associated with suppressed concentration fluctuations can persist up to surprisingly large system sizes, particularly for assemblies involving many strands.¹⁰⁸ Furthermore, it is often no longer straightforward to apply rare-event approaches be-

cause of the problem of defining an appropriate order parameter to drive the formation of multiple copies of a target, or the need to use more replicas for parallel tempering.

We note that, to date, these finite-size effects have generally been neglected.^{53,59–61,69,109–111} However, we have recently developed an analytic approach that allows the bulk probabilities to be estimated from the “single-target” simulations assuming both that the species behave ideally and that other multi-chain aggregated states do not contribute to the thermodynamics.^{93,108,112} Both assumptions are likely to hold well for DNA, firstly, because ideality is a good approximation for the dilute DNA solutions typically used, and secondly, because the targets one usually wants to consider are maximally base-paired, so any alternative misbonded aggregate, although perhaps relevant as a kinetic trap, is likely to have higher free energy than the target and so not be thermodynamically relevant. This procedure is simple to implement, and so we recommend that it should be applied for any calculation of thermodynamic properties from single-target simulations.

4.1.2 Accelerating kinetic measurements. Accelerating kinetic measurements is more difficult than enhancing thermodynamic equilibration, as simulations must be unbiased and transitions must occur at the temperature of interest. In general, methods deal with measuring the flux of trajectories from one local free-energy minimum *A* to an alternative *B* (a quantity analogous to transition rates for non-instantaneous processes), and also generating a representative set of transition paths.

We have found forward flux sampling to be a particularly useful algorithm of this kind.^{113,114} First, the transition is split into several stages by defining interfaces which must be crossed in the transition from *A* to *B*. Simulations are performed in the vicinity of *A* to measure the flux of trajectories across the first interface. States are saved at the crossing point. These states are then used to initiate new trajectories, which are used to measure the probability of reaching the next interface before returning to *A*. This process is repeated for subsequent interfaces. The overall flux from *A* to *B* is then given by the initial flux multiplied by the probability of success at each interface, and the trajectories generated are an unbiased sample from the ensemble of reaction pathways. By splitting one unlikely process into several stages, the overall measurement can be performed more efficiently.

Forward flux sampling is most effective for simple processes which do not involve long-lived metastable intermediates, such as hybridisation of non-repetitive duplexes.⁸⁸ Although the algorithm is still valid when such metastable intermediates exist, it does not effectively accelerate the resolution of the intermediate into state *A* or *B*, and hence this subprocess must be simulated by brute force. In some cases this is practical, but often it is not. This difficulty arises because one

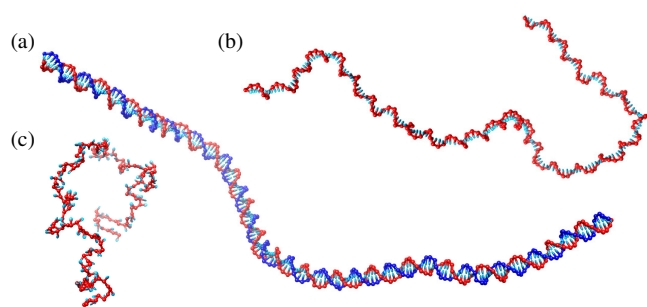


Fig. 2 Typical configurations illustrating the relative flexibility of (a) double-stranded DNA, (b) stacked single-stranded DNA and (c) unstacked single-stranded DNA.

must measure the number of trajectories that return to state *A* at each stage, rather than just the previous interface.

A possible solution to this problem is to explicitly define a state *C* corresponding to the metastable intermediate. One can then use forward flux sampling to measure the flux from *A* to *C*, and then from *C* to *B* and *A*, from which the overall kinetics of the process can be inferred. It is also possible to consider multiple metastable intermediates. Such an approach, which we have used to study internal displacement of repetitive sequences,⁸⁸ relies on the assumption that the system equilibrates in state *C* before it resolves into state *A* or *B*. We note that alternative techniques, such as transition path sampling¹¹⁵ and transition interface sampling,¹¹⁶ face similar difficulties when confronted with metastable intermediates.

5 Results for oxDNA

5.1 Basic DNA biophysics

Fig. 1(b) shows a DNA double helix as represented by oxDNA. By design it has the correct basic structure, e.g. pitch, base-pair rise and radius. However, as already noted, the two grooves are of the same size, whereas in real DNA the major groove is larger. The double helix also exhibits non-trivial features such as propeller twist of the bases that arises in the model due to a competition between optimizing the stacking and hydrogen-bond interactions.⁶⁶

The model also provides a good description of the mechanical properties of DNA. For example, Fig. 2 illustrates the relative persistence lengths of dsDNA and ssDNA, in its stacked and unstacked forms, which in our model are 125 base pairs, and 40 and 3 bases, respectively, in good agreement with available experimental data.⁶⁶ In particular, the relative flexibility of unstacked DNA is important for nanotechnological applications. This flexibility also enables hairpins to form easily.

Perhaps unsurprisingly, given their importance in the fitting procedure, we are able to reproduce well the melting points

and transition widths of DNA duplexes. In order to capture the dependence of the melting point on length, the correct balance between the contributions of stacking and hydrogen-bonding to the stability of the duplex was required. More impressively, the melting points of hairpins are also reproduced well, albeit with a small but constant offset of approximately 2 K. Single-strands in oxDNA also show a broad uncooperative stacking transition, although in this case the experimental nature of this transition is less clear. However, some further information on this transition can be obtained indirectly from the melting temperatures of hairpins, which decrease when the stacking in the loop is sufficiently strong to hinder the bending back of the strand in the loop,¹¹⁷ an effect that our model is able to reproduce.⁸³

One of the features of our model is that it allows the free-energy landscapes associated with various assembly processes to be computed.^{23,28,29,66,83–87} For example, Fig. 3(a) depicts the free-energy profile for the hybridization of a 15-base-pair duplex, where the number of base pairs has been used as the order parameter. There is an initial free-energy barrier associated with the loss of translational entropy on association followed by a linear downhill slope as the number of base pairs increase (the linearity is because the free-energy of forming a base pair is the same for all base pairs in our “average base” parameterization). Notably, the most stable state of this duplex at this temperature is with only 14 of the 15 bases paired. This entropic opening of the duplex at its ends is termed “fraying”, and provides an example of how oxDNA’s behaviour shows deviations from “two-state” behaviour; this contrasts with, for example, SantaLucia’s nearest-neighbour model for predicting oligonucleotide melting temperatures that assumes the two states in equilibrium are a fully-formed duplex and an unstructured single strand with no variation of the states involved with temperature.¹⁶

A simplistic interpretation of this free-energy profile might be to say that the “transition state” to duplex formation involves the formation of a single base pair. However, it is not quite that straightforward, because even when a single base pair forms the rest of both chains normally require considerable conformational change to reach a state where the duplex can zipper up, and so consequently the system is actually much more likely to dissociate.⁸⁸ Similarly, when a duplex is melting and the last base pair breaks, it does not initially feel the extra translational entropy it can gain from being disassociated, and so is still quite likely to zipper back up.

We should note that the mechanism of hybridization that we observe for oxDNA differs strongly from that found by de Pablo and coworkers for their 3SPN.1 DNA model.^{53,54,111,118,119} For oxDNA, hybridization normally occurs by the formation of a nucleus of a few correct base pairs, followed by the “zippering” up of the rest of the duplex as bases are transferred from the relatively unstructured single-

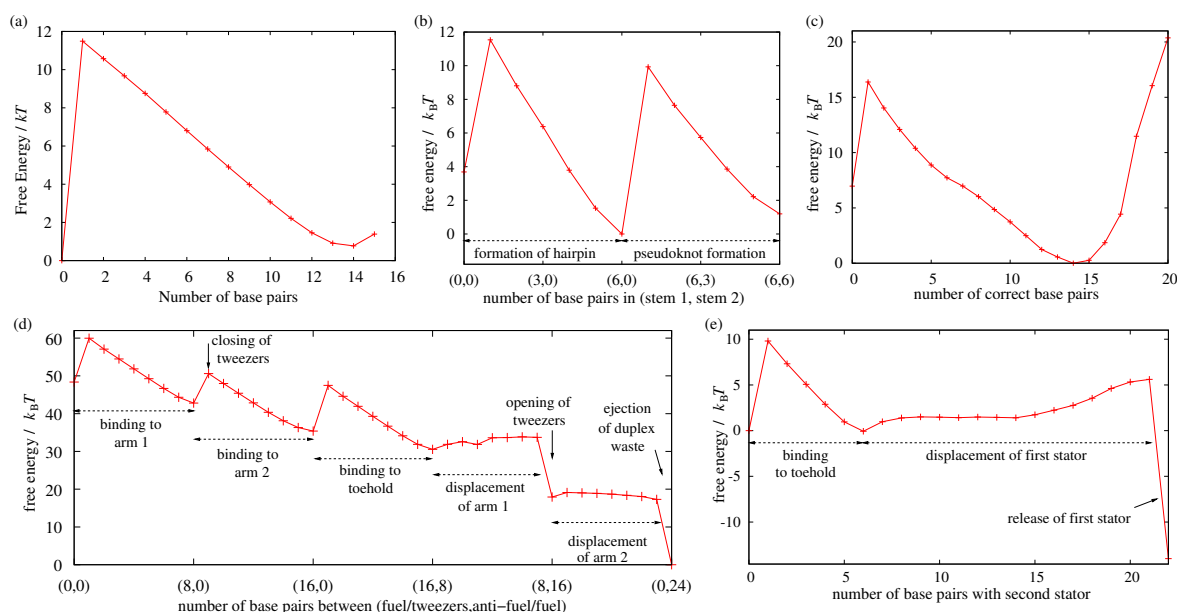


Fig. 3 Free-energy profiles for (a) the formation of a 15 base-pair duplex at 343 K (b) the formation of a pseudoknot at 307 K (Fig. 7(d)), (c) the formation of a kissing complex at 296 K between two hairpins with 20-base complementary loops (Fig. 7(e)), (d) the complete cycle of the DNA nanotweezers illustrated in Fig. 9(a), and (e) one step of the burnt-bridges walker illustrated in Fig. 9(b). All the results are for the average-base parameterization of oxDNA. In (c) the topological frustration leads to a significantly lower free-energy gain from binding than for a duplex with the same number of bases and the most favourable state of the kissing complex has only 14 of the 20 possible base pairs formed.

stranded tails to the growing end of the double-helical section, as illustrated in Fig. 4(a). Alternatively, if a misbonded (i.e. involving base pairs not present in the fully-bonded duplex) nucleus forms first, the fully base-paired duplex can still be achieved by an internal displacement mechanism,⁸⁸ in which the correctly-bonded helical section nucleates and grows at the expense of the misbonded helix. By contrast, for the 3SPN.1 model ssDNA adopts an overly stiff helical geometry, and hybridization can occur from a misbonded duplex by a sliding of the two chains past each other along the helical axis. In our opinion,⁸⁸ this mechanism is an artefact of their overly stiff ssDNA that restricts the way DNA can hybridize. In particular, both the zipper mechanism shown in Fig. 4(a) and the internal displacement mechanism would not be feasible in 3SPN.1. Furthermore, the isotropic nature of the 3SPN.1 base-base attractions and the *ad hoc* sugar-sugar attraction leads to an artificially low barrier to sliding.

To illustrate some of the general principles of self-assembly in DNA systems, in Figure 5(a) we show the hairpin closing times for a sequence that can form misbonded structures which can compete with the formation of the correct hairpin. Like many other self-assembly processes, there is a temperature window where assembly is most efficient. At high temperatures, the hairpin formation rate goes down due to the decreased probability of successfully forming a complete stem

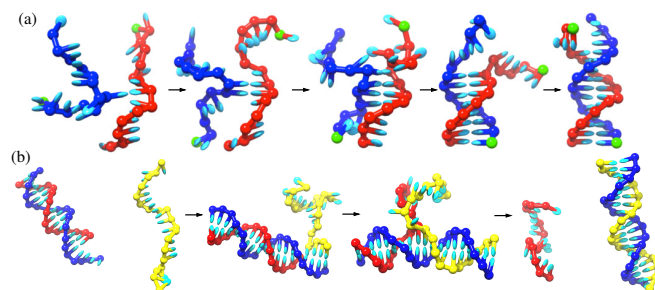


Fig. 4 Example configurations from the pathways for (a) hybridization and (b) toehold-mediated strand displacement.

after the formation of an initial contact. Furthermore, above the melting temperature, the equilibrium fraction of hairpins will decrease rapidly. At low temperatures the hairpin folding times go up due to the formation of misbonded hairpin structures that act as kinetic traps and hinder the formation of the target hairpin. The system can still escape from these misbonded structures by internal displacement processes, but this becomes increasingly difficult as the temperature is decreased. The two most stable misbonded structures are illustrated in Fig. 5(c) and (d). As it happens the more thermodynamically stable of the two is a less effective kinetic trap, because it can rearrange to form the target hairpin relatively easily just by

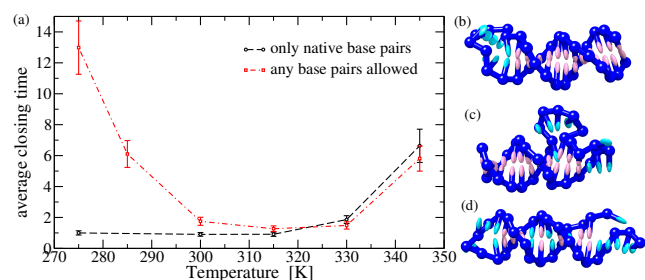


Fig. 5 (a) Folding times for a hairpin as a function of temperature for the sequence TTACATAAAAAGTTTTTTTCTTTTATGTAA. Also represented are folding times when misbonding is prevented by only allowing base pairs in the target hairpin to form. The folding times are given relative to the fastest folding system. (b) The target hairpin. It has a 12-base-pair stem and an 8-base loop, and a melting point of 347 K. (c) The most stable misbonded structure. It has 11 base pairs and a melting temperature of 334 K. (d) The second most stable misbonded structure, which has 7 base pairs and a melting temperature of 314 K. In (b)–(d) the base sites of nucleotides participating in Watson-Crick base pairs are coloured pink to aid visualization of the bonding pattern.

the propagation of the bulge loop to the hairpin loop end of the helical stem. If the formation of misbonded structures is prevented by artificially turning off non-native base pairing, there is no low temperature decrease in the hairpin formation rate.

This example shows some of the typical features of DNA self-assembly. Namely, the existence of an assembly window between T_m (the temperature at which the target structure melts) and T_{misbond} (the temperature at which the most stable misbonded structure melts), where the only structure that is stable with respect to the single-stranded state is the target structure. This is why annealing the system, i.e. cooling it down from a high temperature where all the strands are in the single-stranded state, is such a simple and effective assembly strategy. The system will always pass through the assembly window first and, as long the cooling rate through this window is sufficiently slow, the correct structure will form and the system will not ever get kinetically trapped in misbonded structures. Sequence design can be used to make self-assembly easier by minimizing the stability of misbonded structures, hence making the assembly window wider. In addition, for complex structures that are designed to form hierarchically (i.e. subunits that form at high temperature themselves assemble into larger aggregates at lower temperature) there will be multiple assembly windows, which will be sequentially passed through on annealing.

5.2 DNA under stress

In the cell, DNA is acted upon by a great variety of molecules, and many of these induce some stress on the DNA.^{120,121} For example, protein binding can cause DNA to bend sharply (e.g. the wrapping of DNA around histones in order to package DNA in eukaryotic cells). There are also families of enzymes responsible for manipulating DNA in different ways; e.g. topoisomerases control the supercoiling of DNA, and helicases cause DNA to unwind.

As such, the mechanical properties of DNA are of considerable interest, and the advent of single-molecule techniques, such as optical and magnetic tweezers, to exert forces and torques on individual DNA molecules has led to a wealth of new and impressive data.³ However, such techniques cannot directly detect the microscopic nature of the structural changes induced by the mechanical stress, and so simulations can play an important role in visualizing such structural changes. As noted already, oxDNA was fitted to give a reasonable description of the persistence length of dsDNA. Similarly, the twist modulus of dsDNA 475 fJ fm is at the upper end of the range of values determined from experiment, although this comparison is complicated by the fact that experiments often measure an effective twist modulus that is reduced by thermal bending fluctuations.¹²² However, the stretch modulus of 2100 pN is roughly twice the experimental value.⁶⁶ Note, when fitting the oxDNA model, parameters could not be found that allowed the model to simultaneously reproduce the experimental persistence length and stretch modulus and so we choose for the model to better reproduce the persistence length, as this is more likely to be relevant to the nanotechnological applications.

On pulling on opposite ends of dsDNA, it is found to undergo an overstretching transition to an overstretched form with an approximately 70% increase in length. At room temperature and 500 mM salt this occurs at 67–8 pN.^{123,124} OxDNA undergoes a similar transition at 74 pN, satisfyingly close to the experimental value, and also provides a good description of the temperature dependence of the overstretching force.⁸⁶ The nature of the overstretched state has been the subject of considerable debate, the two main proposals being force-induced melting by unpeeling¹²⁵ and a transition to an overstretched duplex state termed “S-DNA”,¹²⁶ but with recent evidence suggesting that both mechanisms can occur depending on solution conditions and temperature.^{127,128} Interestingly, for our model we only ever see overstretching by unpeeling (Fig. 6(a)). This result strongly suggests that S-DNA is unlikely to be an unstacked but base-paired duplex, one of the proposed structures for S-DNA,¹²⁶ as our model has a good representation of stacking and base pairing and would be expected to reproduce such a structure if it were relevant to overstretching. Instead, S-DNA is likely to have a more exotic

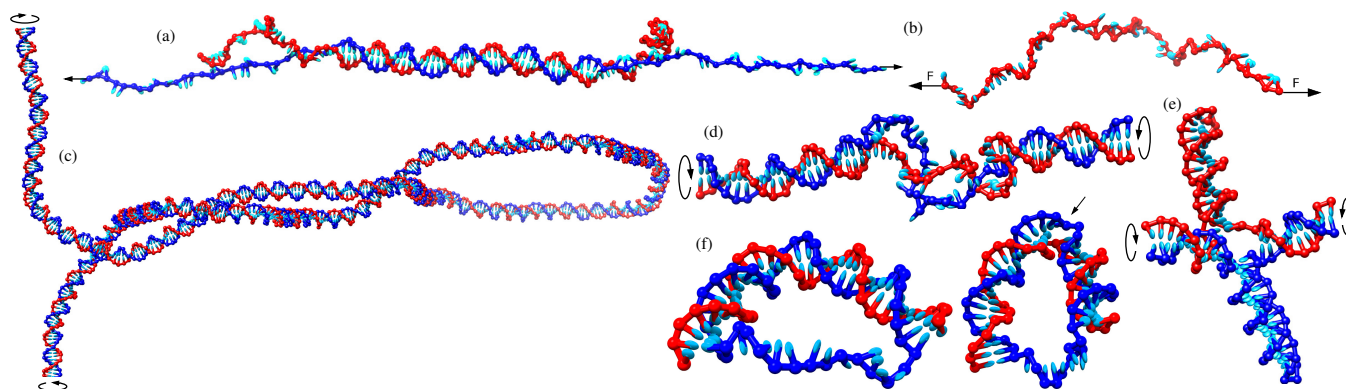


Fig. 6 Response of DNA to mechanical stress. (a) The unpeeling of duplex DNA and (b) the stretching of ssDNA under tension, and the formation of (c) a plectoneme, (d) a bubble and (e) a cruciform for a sequence with an inverted repeat, all under negative supercoiling. (f) Two structures formed by the hybridization of a 35-base linear strand with a 45-base circular strand. For the left-hand structure, the double helical section is homogeneously bent, whereas the other contains a “kink” (labelled by an arrow) where two base pairs are broken.

structure.

The stretching of ssDNA has been less studied, but has a number of interesting features. For a random sequence, at low force and high salt, it will form some secondary structure, e.g. misbonded hairpins. The breaking of this secondary structure leads to a feature in the force-extension curve (below 10 pN for 0.5 M salt).¹²⁹ However, such secondary structure can be prevented by choosing sequences possessing no complementary bases. If such sequences have strong stacking, features in the force-extensions curves associated with force-induced unstacking can be observed.¹³⁰ However, our model suggests that this transition is not associated with complete loss of stacking, but rather with a shortening of the length of runs of stacked bases.⁸³ For example, even in the force-bearing single-stranded sections of the duplex undergoing unpeeling in Fig. 6(a) one can still see short runs of three or four stacked bases, where the backbone is able to align itself along the axis of force.

DNA's response to applied twist is dependent on its sign (undertwist is termed negative supercoiling, and overtwist positive supercoiling), the presence of tension and sequence. At low force for long DNAs, the molecule is able to absorb the twist by writhing to produce plectonemes,¹³¹ such as in Fig. 6(c). For short DNAs or at higher forces, negative supercoiling can be absorbed in other ways, for example by the formation of a bubble, as in Fig. 6(d). One noteworthy feature of such bubbles in our model is that the two strands twist around each other in the opposite sense to the DNA double helix. This is so that the bubble absorbs as much of the negative twist as possible, thus minimizing the number of base pairs that need to be broken. For sequences possessing an inverted repeat (a stretch of dsDNA where the sequence reads the same on either strand) the molecule can absorb the negative supercoiling by forming a double hairpin structure called a “cruciform” (Fig.

6(e)). Using oxDNA, we have observed that there is a cooperative but asynchronous mechanism for the formation of the hairpin arms of the cruciform. The nucleation of the first hairpin from a bubble that has diffused to the centre of the inverted repeat makes the nucleation of the second arm much more likely.⁸⁵

There has also been much interest in the bending fluctuations of dsDNA.¹³² Cyclization rates for large DNAs agree well with the predictions of the wormlike-chain model.¹³³ However, shorter duplexes (of the order of the persistence length and less) show enhanced cyclization rates compared to this model.^{6,134} It has been suggested that these enhanced bending fluctuations are due to kinking at sites where transient bubbles form.^{135,136}

Bending of DNA can also be induced by hybridization of a circular strand with a linear strand to give a duplex where the two ends of the double helix are connected by a single strand.^{137,138} As this single-stranded section becomes shorter, the structure of the double-stranded section changes from being homogeneously bent to developing a kink due to the formation of a bubble roughly in the centre of the double-stranded section. For the example illustrated in Fig. 6(f), both the bent and the kinked structures are possible for oxDNA. For the homogeneously bent case, the tension in the single-stranded section enhances fraying of the double helix. However, this tension is significantly reduced on kinking.

The mechanical properties of DNA nanotechnology systems are also beginning to be probed. For example, single-molecule measurements of the persistence lengths and torsional moduli of DNA origami multi-helix bundles have recently been obtained using magnetic¹³⁹ and optical¹⁴⁰ tweezers, with the extreme rigidities of these origamis making them particularly suitable for use as linkers in tweezer experiments.¹⁴⁰

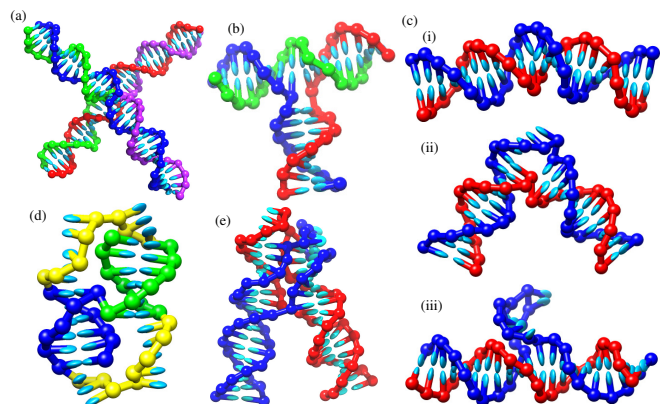


Fig. 7 Simple DNA motifs: (a) a 4-way junction; (b) a 3-way junction; (c) duplexes with a bulge loop of (i) 2, (ii) 5 or (iii) 9 bases; (d) a pseudoknot and (e) kissing hairpins.

5.3 Structural DNA Nanotechnology

As DNA nanotechnological structures basically consist of double helical sections connected by junctions and single-stranded sections, it will be useful to consider the properties of such junctions. The most common junction is a four-way junction. At the intermediate to high salt typically used in DNA nanotechnology, the junction does not adopt an open geometry, but instead adopts a stacked structure, where there is coaxial stacking between the arms at the junction site, so that there are two (quasi)-continuous helices passing through the junction. These helices prefer not to be parallel, and adopt an “X-like” configuration. Fig. 7(a) shows such a junction for oxDNA. There are two possibilities for the chiral twist of the two helices with respect to each other. In our model the junction preferentially adopts the left-handed form, and although this form has been seen experimentally,¹⁴¹ the norm is for the junctions to be right-handed.¹⁴² Presumably, oxDNA has insufficient structural detail to reproduce this local preference. However, this deficiency is likely to be not too detrimental to the study of the nanotechnological systems, since the junctions are usually not free to adopt their preferred local structure, but are constrained by other parts of the nanostructure. In particular, the helices are often roughly anti-parallel, as, for example, in DNA origami. However, the internal stresses that result would be expected to be of the incorrect sign with respect to the twisting of the helices at the junctions. We are currently exploring these rather subtle effects further in order to understand how they affect large DNA nanostructures.

3-way junctions, although somewhat less common, are another basic motif that is used in DNA nanotechnology,⁷ for example at the corners of a DNA nanotetrahedron.¹⁴³ The two basic proposed types of structure for the junction are a T-shaped geometry, where two of the arms coaxially stack at the junction, or a geometry with the arms at approximately 120°

and an open arrangement at the junction without any coaxial stacking.¹⁴² The competition between these structures depends on solution conditions and whether there is a bulge loop at the junction — this provides the structure with more flexibility making coaxial stacking easier. The T-shaped geometry that we find for oxDNA is illustrated in Fig. 7(b); the identity of the two arms that are coaxially stacked was found to switch frequently.

Another common motif that can be used to help control the structure of DNA assemblies is a bulge loop where one strand has a section of extra non-complementary bases. Some example structures for duplexes with bulges are shown in Fig. 7(c) with the structures reflecting the nature of the stacking in the bulge region. If the bases on either side of the bulge coaxially stack as in Fig. 7(c)(iii) the duplex is relatively unperturbed, whereas if the stacking is broken between bases on the non-bulge-containing strand the bulge acts as a relatively flexible hinge allowing larger angular deviations between the two double helical sections (Fig. 7(c)(ii)). In between these two cases is Fig. 7(c)(i) where the bases in the bulge stack with those of the duplex and there are no breaks in the stacking in the other strand, and a small bend is induced.

The connectivity of the double helical sections can also sometimes lead to pseudoknotted configurations.^{81,88} The example pseudoknot illustrated in Fig. 7(d) is effectively a hairpin with a tail that binds to the loop to form a second helical section. The thermodynamics of this system are interesting since, although the two helices in this example are equivalent, the free-energy profile in Fig. 3(b) shows that going from either of the two possible hairpins to the pseudoknot by forming a second helical section is thermodynamically less favourable than forming the hairpin. As a consequence, as the system is cooled down, the system will form a hairpin at 320 K but only at 299 K does the pseudoknot become most stable. It is also noteworthy that secondary structure thermodynamic models find pseudoknots particularly challenging,¹⁴⁴ again illustrating the extra insights available from models with a full three-dimensional representation of the structure.

One of the first larger-scale self-assembled DNA nanostructures were the two-dimensional arrays of double-crossover tiles¹⁴⁵ produced by Winfree *et al.*¹⁴⁶ An individual double-crossover tile is illustrated in Fig. 8(a), and a small 4 × 4 array of such motifs in Fig. 8(b). Their self-assembly is hierarchical with the double-crossover tiles forming at higher temperature, and then these further self-assembling into arrays mediated by interactions between the tiles’ short single-stranded “sticky ends”. The arrays are usually visualized after deposition on a surface, and are then of course flat. However, their structure and flexibility in solution is less clear, although the tendency of some tile types to form tubes suggests a preference for these sheets to curved.¹⁴⁷ Particularly noticeable in the structure in Fig. 8(b) are the gaps that open up between adjacent tiles, and

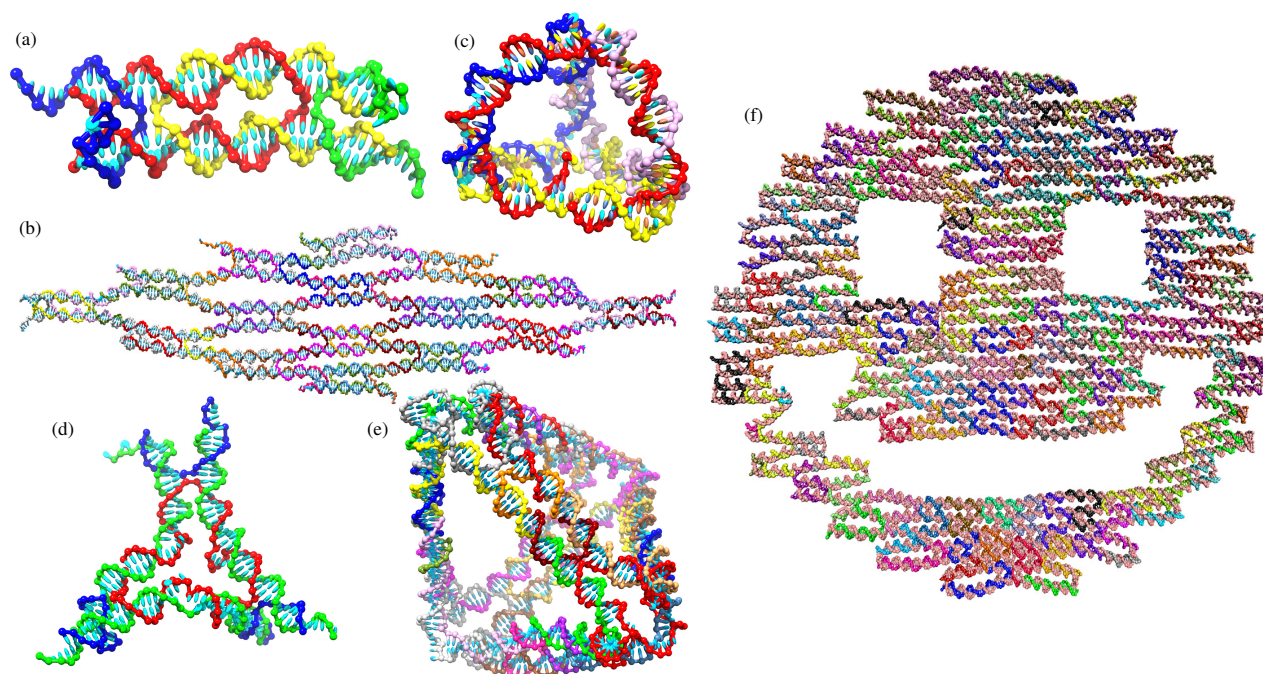


Fig. 8 Examples of DNA nanostructures: (a) a double crossover tile, and (b) a 4×4 array of such tiles adsorbed on a surface, (c) a nanotetrahedron, (d) a three-armed star motif and (e) a tetrahedron made up of four such motifs, and (f) a “smiley-face” DNA origami made up of 6196 base pairs adsorbed on a surface.

this compares well to the “rectangular checkerboard” pattern typically seen in experimental micrographs, where the gaps are of the order of half of the width of the tile.¹⁴⁷ Our structure confirms that these gaps are the result of bending of the helices, particularly near to the junctions, but without significant loss of stacking at the junctions or strand ends.

We have also begun to explore the assembly of the tiles into arrays by dynamics simulations. As when we considered the hairpin, there will be a temperature window just below the melting point of the crystals where the only stable product is the correctly-formed array. However, in this temperature window there is a “nucleation” free-energy barrier to assembly, and so hysteresis is experimentally observed between assembly and disassembly of the arrays.^{148,149} In particular, tiles bound by a single sticky-end to the rest of the array will not be stable, and have a significant likelihood of dissociating, before one of its other sticky-ends either find its partner in the aggregate or binds to a new tile joining the aggregate. Consequently, simulating assembly in this regime is challenging, because many binding events do not lead to growth, and nucleation of the initial array is especially difficult. In particular, a dimer will not be stable, and only once a nucleus of four tiles is formed can each tile be bound by two of its sticky ends. It is far easier to simulate assembly for this system in a lower temperature range where every sticky-ended associa-

tion is stable, and so there is no nucleation free-energy barrier. However, this can lead to imperfect structures, because, for example, before all the sticky ends of a tile find their correct binding partner in an aggregate, another tile can come in and stably bind to one of those partners — this effect is also exacerbated by the higher concentrations that one typically uses in simulations to reduce diffusion times. Forward-flux sampling provides a potential means to operate within the “nucleation” regime by sampling the successive addition of tiles to a growing aggregate.

DNA strands can also be designed to form finite structures of a particular size and shape. Fig. 8(c) illustrates a nanotetrahedron designed by the Turberfield group to assemble from four strands, where each strand runs around one face of the structure.¹⁴³ In this case, the structure is designed to form hierarchically, with dimers involving the two edges with no nicks forming first at higher temperature. These two types of dimers will then themselves dimerize with each other to form the tetrahedron as the system is cooled further.

A different route to polyhedra is through the multi-arm motifs developed particularly in the group of Chengde Mao.^{150–154} An example three-arm tile is shown in Fig. 8(d). The arms have a very similar structure to one half of the double-crossover tiles, including single-stranded sticky ends that allow the motifs to assemble into higher-order structures.

To impart flexibility to the motifs, the red strand in Fig. 8(d) has three bulge loops at the centre of the motif. It is noticeable that for the configuration in Fig. 8(d), there is a coaxial stacking across the bulge in one case leading to a single homogeneously bent double helix bridging two of the arms, whereas the stacking is broken at the other two bulges giving rise to a kink. It is also noticeable that the two helices in an arm are not necessarily parallel, reflecting the preference for the coaxially stacked helices in a four-way junction to be twisted with respect to each other.

The interactions and assembly products of these multi-arm motifs can be controlled in a number of ways, including through the length and number of the arms, the number of bases in the bulges and the strand concentration. If the length of two arms bound to each other is an integer multiple of complete helical turns, then the two motifs have the same orientation, and so any non-planarity of the motifs can lead to a build up of curvature, making closed structures more likely. By contrast if the length is an odd multiple of half helical turns, the motifs will have opposite orientation, and any non-planarity will be cancelled out, making planar crystalline arrays more likely. This approach has been used to create octahedra¹⁵⁰ and icosahedra¹⁵¹ from 4- and 5-arm motifs respectively, and honeycomb,¹⁵² square¹⁵³ and hexagonal¹⁵⁴ lattices from 3-, 4- and 6-arm motifs, respectively.

The case of closed structures with 3-arm motifs is particularly interesting,¹⁵⁵ as the structures that result reflect the relative time scales for internal structural fluctuations in a growing aggregate that allow closed polygons to form and for addition of new motifs to the growing structure. When the bulges are five bases long and the motifs particularly flexible, tetrahedra form — it is noticeable from the tetrahedron in Fig. 8(e) how the unpaired bases in the bulge loops are stretched out to bridge the helices in the different arms and allow the sharp back-bending at the vertices of the tetrahedron. When the bulges are shorter and the motifs less flexible, larger structures such as dodecahedra and truncated icosahedra can form, with the larger structure being favoured by higher concentration, as this increases the rate of addition of new motifs to a growing structure. We intend to use oxDNA to probe how the length of the bulge loops affects the rate of polygon closing using forward-flux sampling.

Comparing the two tetrahedra in Fig. 8(c) and (e), it is noticeable that the edges of the larger tetrahedron are significantly straighter; the presence of two linked helices along each edge leads to a substantial enhancement in their stiffness.

A significant advance in the repertoire of nanostructures that could be reliably formed was through the development of the DNA origami technique by Rothemund,¹⁵⁶ in which a long single-stranded viral “scaffold” DNA is folded up into a structure made of linked parallel double helices by the addition of lots of short “staple” strands. The iconic smiley-

face origami (Fig. 8(f)) illustrated that virtually any arbitrary two-dimensional shape could be formed. The technique was then further extended to allow three-dimensional¹⁵⁷ and curved^{158,159} structures to form.

The origami illustrated in Fig. 8(f) shows that our model is able to capture one of the most basic structural features of DNA origami, namely the “weave” pattern where the spacing between the helices opens up in between the junctions. However, the structure of origamis are virtually always probed when bound to a surface, and much less is known experimentally about their structure in solution, although the CanDo package can provide a useful guide.^{79,80} We find that two-dimensional origamis can have quite large structural fluctuations, and need not be planar on average. As well as the weave pattern, we find that the origamis can be “corrugated” in the plane perpendicular to the origami, as has also been seen in the recent cryo-EM structure of a three-dimensional origami.¹⁶⁰ The original origami of Rothemund are almost certainly twisted in solution, because the design requires the helical repeat length to be $10\frac{2}{3}$ base pairs, and so the mismatch between that and DNA’s natural helical repeat of about 10.5 bases per turn leads to twist, as has been illustrated for long three-dimensional origami ribbons.¹⁵⁸ For oxDNA, we find that the overall twist also reflects the effects of nicks, and junctions on the local twist angle at the corresponding base-pair steps.

We note that a programme to take output from cadnano,¹⁶¹ a commonly used DNA origami design tool, and turn it into an input file for oxDNA is available at the oxDNA website.⁹²

Experiments have shown that there is hysteresis between the self-assembly and melting of DNA origamis, that the self-assembly can occur over a relatively narrow temperature window, and that if the system is quenched below this assembly window the quality of the resulting origami decreases.^{162,163} These results suggest that near the melting temperature there is a nucleation free-energy barrier to origami formation; this barrier probably reflects the decrease in conformational entropy of the remaining single-stranded sections of the scaffold strand as staple strands bind making the binding of each successive staple more thermodynamically favourable. We have begun to use oxDNA to explore the self-assembly of small DNA origamis, but, similar to the assembly of double-crossover tiles, direct simulations of assembly have to be in a temperature regime where the assembly process is downhill in order to occur in a reasonable amount of computer time. However, we then see kinetic trapping in configurations where two copies of the same staple have both bound to the scaffold, thus blocking either staple strand fully bonding to the scaffold. In order to avoid such problems, one would need to work at a temperature where the binding of the first domain of each staple is unfavourable, so that singly-bound blocking strands would melt away, but this regime would require rare-

event techniques, such as forward-flux sampling to be used.

5.4 Active DNA nanotechnology

In DNA nanodevices, activity is often achieved through what are termed strand-exchange or displacement reactions, in which an “incumbent” strand that is partially complementary to a “substrate” strand is replaced by an “invading” strand that is able to form more base pairs with the substrate. Figure 4(b) illustrates this process. Interestingly, the rate of displacement is found to initially increase exponentially with the length of the toehold (the number of unpaired bases in the substrate-incumbent complex) before plateauing at 5 or 6 bases.⁵ In contrast to simple kinetics schemes based on nearest-neighbour thermodynamic models, oxDNA is able to quantitatively reproduce the $10^{6.5}$ -fold acceleration of the rates with increasing toehold length, because it can capture the free-energetic effects of the overcrowding at the junction between the invading and incumbent strands and the dynamical consequences of the greater structural rearrangement necessary for displacement to progress compared to toehold melting.⁸⁷

One of the first nanodevices to use displacement was the DNA nanotweezers developed by Yurke *et al.*,¹³ an example of which is illustrated in Fig. 9(a).²³ The nanotweezers cycle between “open” and “closed” configurations through the “clocked” addition of complementary single strands. Using oxDNA, we have been able to characterize the free-energy landscape associated with a complete cycle of the nanotweezers (Fig. 3(d)). As the two strands would preferentially hybridize with each other rather than open or close the tweezers, their addition is necessarily sequential in time. One way round this restriction is to prepare these “fuel” strands in a form that prevents them from hybridizing directly, but where the nanodevice can extract work by catalysing their hybridization. This can be achieved if the strands are designed to form hairpins.^{164,165} Although the hairpin loops can start to hybridize, the topology prevents full hybridization (Fig. 3(c))—the linking number of the two loops must be conserved so for every time the two loops wrap around each other to form the B-DNA double helix, they must also wrap around each other in the opposite sense. The structure of the “kissing complex” formed by two such hairpins is illustrated in Fig. 7(e). Perhaps contrary to expectation, oxDNA suggests that it is favourable for the loops to form two double-helical sections adjacent to each hairpin stem rather than a single hybridized region.⁸⁴

An alternative approach to introduce dynamical activity is to use a modified restriction enzyme to nick one of the strands in a double helical section, such that it is favourable for at least one of the two fragments to dehybridize. This approach has been used to create a “burnt-bridges” motor that walks along a track of single-stranded stators attached to some substrate (be

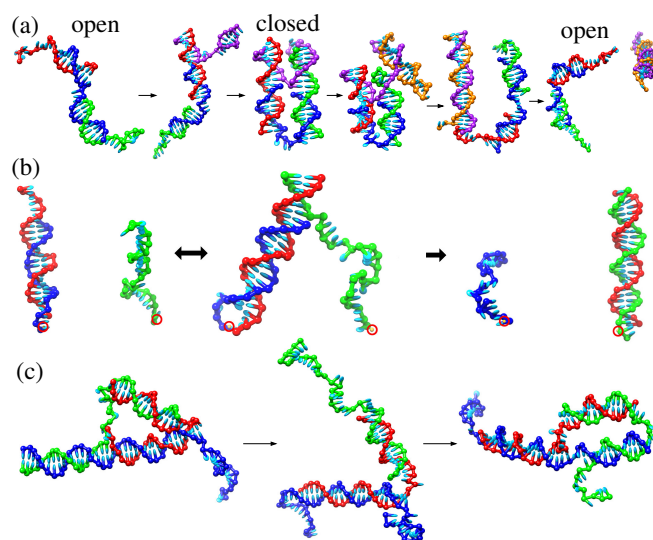


Fig. 9 Examples of DNA nanodevices: (a) nanotweezers transforming between their open and closed states in response to the addition of fuel (purple) and anti-fuel (gold) strands, (b) a burnt-bridges walker moving between two stators, and (c) a two-footed DNA walker taking a step along a single-stranded track (blue). In (b) the red circles indicate the points of attachment to the substrate.

it a single double helix¹⁶⁶ or a DNA origami¹⁵) and is illustrated in Fig. 9(b). The enzyme nicks the stator-cargo complex to reveal a toehold that allows the cargo to be transferred to the next stator through a displacement reaction. Thus, the track is modified as the motor progresses, hence the “burnt-bridges” name. We have used our model to probe the free-energy landscapes for the displacement process, and particularly the effect of stator separation.²⁹ As the stators become increasingly separated, there is an increasing free-energy barrier to displacement because of the tension that builds up in the intermediate as it is stretched between the binding sites of the two stators (Fig. 3(e)). This effect is potentially useful for tracks where one wants the cargo to select one of two paths,¹⁶⁷ because as the probability of displacement being successful diminishes (due to the large free-energy barrier), the relative rates of taking either path become more sensitive to the toehold binding strengths.

We have also modelled a prototype two-footed walker developed in the Turberfield group that walks along a single-stranded track. There are two versions of the walker, one that makes use of nicking enzymes¹⁶⁸ and one that uses hairpins as fuel,¹⁶⁹ the former being illustrated in Fig. 9(c). The motor is designed so that the two-feet overlap, and hence expose a toehold that allows either of the feet to be selectively detached from the track by displacement using an appropriate fuel strand. Therefore, in the presence of the fuel that lifts the

back foot, the walker should walk forwards. Our modelling suggests that there is also a potential bias, stemming from the geometry of the walker, for the lifted foot to reattach in the forward direction, especially when moderate tension is applied to the track.²⁸ However, we also found that the walker had a tendency to “overstep”, and attach to a binding site not adjacent to the stationary foot, which is a potential problem because the feet then no longer overlap, making it harder for the feet to be lifted by the fuel strands and eliminating the intended bias for lifting the back foot. Applying moderate tension to the track was found to make recover from the overstepped state easier. This study probably provides the first example of optimization strategies for a DNA nanodevice being suggested based on simulations.²⁸

6 Conclusions

Molecular simulation has potentially much to offer the fields of DNA nanotechnology and DNA biophysics. However, for this promise to be realized, accurate and robust coarse-grained models of DNA are required in order to address the potentially long time and length scales involved. After a relatively late start compared to coarse-grained simulations of biomolecules such as proteins and lipids, coarse-grained modelling of DNA has received considerable interest over the last few years, with many possible models now available in the literature and an increasing number of applications using these models. However, for quite a number of these models, their fundamental behaviour has not been sufficiently tested to be confident of their ability to capture a wide range of phenomena. Furthermore, many models involve choices concerning the form of their interactions that limit their ability to exhibit the physical properties that are most relevant to applications to the self-assembly processes associated with DNA nanotechnology, namely a realistic description of the structure, thermodynamics and mechanics of both dsDNA and ssDNA.

By contrast, the oxDNA model, which was developed by the authors and is the focus of the second-half of this perspective article, was specifically designed to capture the biophysical processes that are essential to self-assembling DNA nanotechnology. As well as accurately reproducing the structural, thermodynamic and mechanical properties that were involved in the fitting process, oxDNA can be simulated on the diffusive time scales relevant to self-assembly processes, even for systems containing thousands of nucleotides. Furthermore, the power of the model is exemplified by the quantitative reproduction of phenomena to which the model was not fitted, most notably the kinetics of toehold-mediated strand-exchange and the overstretching force of duplex DNA. Its utility is further illustrated by the wide range of examples to which the model has been applied, as outlined in Section 5. It is able to provide significant physical insight into fundamental dynamic

processes involving DNA, such as hybridization, strand exchange and hairpin formation, and the response of DNA to mechanical stress, be it stretching, twisting or bending, as well as providing an excellent description of the structural properties of DNA nanosystems. Moreover, its ability to capture non-trivial geometric features of nanodevices that are not accessible to secondary-structure thermodynamic models gives it a potentially important role in guiding the design of DNA nanodevices, as has been illustrated for a prototype two-footed DNA walker. OxDNA has also been used to probe the liquid-crystalline phase behaviour of concentrated solutions of short DNA duplexes.⁸⁹

Having illustrated oxDNA's achievements, it is also important to be open about its deficiencies and limitations. Probably the main limitation is that the model is fitted to one relatively high salt concentration (0.5 M), albeit one that is typical of the high ionic strengths usually used in DNA nanotechnology. Furthermore, the sequence dependence in the model is limited to just the thermodynamic properties. Thus, there is no sequence-dependent elastic behaviour, which is probably less important for nanotechnology, but may be relatively more important for applications in biology. Additionally, the only form of base pairing allowed is Watson-Crick, and so alternative forms of DNA, such as G-quadruplexes and triple-stranded DNA, cannot be modelled. The symmetric nature of the oxDNA helix is a further simplification, but one that we are currently addressing due to its potential structural effects on DNA nanostructures. Finally, the property most relevant to DNA nanotechnology that oxDNA is not able to reproduce is probably the structure of four-way junctions for which oxDNA has a preference for the isomer with the opposite chirality to that usually observed experimentally.

In summary, coarse-grained modelling of DNA has now reached an exciting stage. Models are now available that allow DNA nanosystems to be systematically and accurately probed, thus opening up the field of DNA nanotechnology to molecular simulations. Even complete DNA origamis with over ten thousand nucleotides can be structurally characterized. Probing the self-assembly mechanisms of such large nanostructures is still a real challenge for simulations because of the long time scales involved, but it is envisaged that through the combination of GPU computing and advanced rare-event techniques simulations of the self-assembly of structures with of the order of a thousand nucleotides will soon be possible.

Acknowledgements

The authors are grateful to the Engineering and Physical Sciences Research Council, University College (Oxford) and the German Academic Exchange Service (DAAD) for financial support. PS is grateful for the award of a Scatcherd European Scholarship and RMH for a National Science Foundation

References

- 1 V. Linko and H. Dietz, *Curr. Opin. Biotechnol.*, 2013.
- 2 J. Bath and A. J. Turberfield, *Nature Nanotechnology*, 2007, **2**, 275–284.
- 3 Z. D. Bryant, F. C. Oberstrass and A. Basu, *Curr. Opin. Struct. Biol.*, 2012, **84**, 304–312.
- 4 S. Forth, M. Y. Sheinin, J. Inman and M. D. Wang, *Ann. Rev. Biochem.*, 2013, **42**, 583–604.
- 5 D. Y. Zhang and E. Winfree, *J. Am. Chem. Soc.*, 2009, **131**, 17303–17314.
- 6 R. Vafabakhsh and T. Ha, *Science*, 2012, **337**, 1097–1101.
- 7 N. C. Seeman, *J. Theor. Biol.*, 1982, **99**, 237–247.
- 8 A. Kuzyk, R. Schreiber, Z. Fan, G. Pardatscher, E.-M. Roller, A. Högele, F. C. Simmel, A. O. Govorov and T. Liedl, *Nature*, 2012, **483**, 311–314.
- 9 M. Langecker, V. Arnaut, T. G. Martin, J. List, S. Renner, M. Mayer, H. Dietz and F. C. Simmel, *Science*, 2012, **338**, 932–936.
- 10 S. M. Douglas, I. Bachelet and G. M. Church, *Science*, 2012, **335**, 831–834.
- 11 S. Modi, M. G. Swetha, D. Goswami, G. D. Gupta, S. Mayor and Y. Krishnan, *Nature Nanotechnology*, 2009, **4**, 325–330.
- 12 M. You, Y. Chen, X. Zhang, H. Liu, R. Wang, K. R. Williams and W. Tan, *Angew. Chem. Int. Ed.*, 2012, **51**, 2457–2460.
- 13 B. Yurke, A. J. Turberfield, A. P. Mills Jr, F. C. Simmel and J. L. Neumann, *Nature*, 2000, **406**, 605–608.
- 14 E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf and J. Kjems, *Nature*, 2009, **459**, 73–76.
- 15 S. F. Wickham, M. Endo, Y. Katsuda, K. Hidaka, J. Bath, H. Sugiyama and A. J. Turberfield, *Nature Nanotechnology*, 2011, **6**, 166–169.
- 16 J. SantaLucia, Jr. and D. Hicks, *Annu. Rev. Biophys. Biomol. Struct.*, 2004, **33**, 415–40.
- 17 S. Takada, *Curr. Opin. Struct. Biol.*, 2012, **22**, 130–137.
- 18 M. G. Saunders and G. A. Voth, *Curr. Opin. Struct. Biol.*, 2012, **22**, 245406.
- 19 W. Shinoda, R. DeVane and M. L. Klein, *Curr. Opin. Struct. Biol.*, 2012, **22**, 175–186.
- 20 J. J. de Pablo, *Ann. Rev. Phys. Chem.*, 2011, **62**, 555–574.
- 21 F. Lankaš, *Innovations in biomolecular modelling and simulation*, Royal Society of Chemistry, 2012, vol. 2, pp. 3–32.
- 22 D. Potoyan, A. Savelyev and G. A. Papoian, *WIREs: Comp. Mol. Sci.*, 2013, **3**, 69–83.
- 23 T. E. Ouldridge, A. A. Louis and J. P. K. Doye, *Phys. Rev. Lett.*, 2010, **104**, 178101.
- 24 C. Bustamante, J. F. Marko, E. D. Siggia and S. B. Smith, *Science*, 1994, **265**, 1599–1600.
- 25 J. P. Peters and L. J. Maher III, *Quart. Rev. Biophys.*, 2010, **43**, 23–63.
- 26 C. A. Laughton and S. A. Harris, *WIREs: Comp. Mol. Sci.*, 2011, **1**, 590–600.
- 27 A. Pérez, F. J. Luque and M. Orozco, *Acc. Chem. Res.*, 2012, **45**, 196–205.
- 28 T. E. Ouldridge, R. L. Hoare, A. A. Louis, J. P. K. Doye, J. Bath and A. J. Turberfield, *ACS NANO*, 2013, **7**, 2479–2490.
- 29 P. Šulc, T. E. Ouldridge, F. Romano, J. P. K. Doye and A. A. Louis, *Natural Computing*, accepted; arXiv:1212.4536.
- 30 R. C. DeMille, T. E. Cheatham III and V. Molinero, *J. Phys. Chem. B*, 2011, **115**, 132–142.
- 31 B. Mergell, M. R. Ejtehadi and R. Everaers, *Phys. Rev. E*, 2003, **68**, 021911.
- 32 H. L. Tepper and G. A. Voth, *J. Chem. Phys.*, 2005, **122**, 124906.
- 33 N. B. Becker and R. Everaers, *Phys. Rev. E*, 2007, **76**, 021923.
- 34 N. B. Becker and R. Everaers, *J. Chem. Phys.*, 2009, **130**, 135102.
- 35 A. Savelyev and G. A. Papoian, *Biophys. J.*, 2009, **96**, 4044–4052.
- 36 A. Savelyev and G. A. Papoian, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 20340–20345.
- 37 O. Gonzalez, D. Petkevičiute and J. H. Maddocks, *J. Chem. Phys.*, 2012, **138**, 055102.
- 38 F. Lankaš, O. Gonzalez, L. M. Heffler, G. Stoll, M. Moakher and J. H. Maddocks, *Phys. Chem. Chem. Phys.*, 2009, **11**, 10565–10588.
- 39 M. Peyrard, *Nonlinearity*, 2004, **17**, R1–R40.
- 40 M. Sales-Pardo, R. Guimera, A. A. Moreira, J. Widom and L. A. N. Amaral, *Phys. Rev. E*, 2005, **71**, 051902.
- 41 F. W. Starr and F. Sciortino, *J. Phys.: Condens. Matter*, 2006, **18**, L347.
- 42 J. C. Araque, A. Z. Panagiotopoulos and M. A. Robert, *J. Chem. Phys.*, 2011, **134**, 165103.
- 43 C. Svaneborg, *Comp. Phys. Comm.*, 2012, **183**, 1793–1802.
- 44 C. Svaneborg, H. Fellermann and S. Rasmussen, *Lecture Notes in Computer Science*, 2012, **7433**, 123–134.
- 45 T. E. Ouldridge, I. G. Johnston, A. A. Louis and J. P. K. Doye, *J. Chem. Phys.*, 2009, **130**, 065101.
- 46 C. Svaneborg, *Natural Computing*, arxiv:1210.6156.
- 47 W. Dei, S. K. Kumar and F. W. Starr, *Soft Matter*, 2010, **6**, 6130–6135.
- 48 K. Drukker and G. C. Schatz, *J. Phys. Chem. B*, 2000, **104**, 6108–6111.
- 49 K. Drukker and G. C. Wu, G. Schatz, *J. Chem. Phys.*, 2001, **114**, 579–590.
- 50 S. P. Mielke, N. Gronbech-Jensen, V. V. Krishnan, W. H. Fink and C. J. Benham, *J. Chem. Phys.*, 2005, **123**, 124911.
- 51 T. A. Knotts IV, N. Rathore, D. C. Schwartz and J. J. de Pablo, *J. Chem. Phys.*, 2007, **126**, 084901.
- 52 F. Trovato and V. Tozzini, *J. Phys. Chem. B*, 2008, **112**, 13197–13200.
- 53 E. J. Sambriski, D. C. Schwartz and J. J. de Pablo, *Biophys. J.*, 2009, **96**, 1675–1690.
- 54 E. J. Sambriski, V. Ortiz and J. J. de Pablo, *J. Phys.: Condens. Matter*, 2009, **21**, 034105.
- 55 M. Kenward and K. D. Dorfman, *J. Chem. Phys.*, 2009, **130**, 095101.
- 56 A. Morris-Andrews, J. Rottler and S. S. Plotkin, *J. Chem. Phys.*, 2010, **132**, 035105.
- 57 K. Doi, T. Haga, H. Shintaku and S. Kawano, *Phil. Trans. R. Soc. A*, 2010, **368**, 2615–2628.
- 58 P. D. Dans, A. Zeida, M. R. Machado and S. Pantano, *J. Chem. Theory Comput.*, 2010, **6**, 1711–1725.
- 59 N. B. Tito and J. M. Stubbs, *Chem. Phys. Lett.*, 2010, **485**, 354–359.
- 60 J. H. Allen, E. T. Schoch and J. M. Stubbs, *J. Phys. Chem. B*, 2011, **115**, 1720–1726.
- 61 G. S. Freeman, D. M. Hinckley and J. J. de Pablo, *J. Chem. Phys.*, 2011, **135**, 165104.
- 62 M. C. Linak, R. Tourdot and K. D. Dorfman, *J. Chem. Phys.*, 2011, **135**, 205102.
- 63 I. P. Kikot, A. V. Savin, E. A. Zubova, M. A. Mazo, E. B. Gusarova, L. I. Manevitch and A. V. Onufriev, *Biophysics*, 2011, **56**, 387–392.
- 64 A. V. Savin, M. A. Mazo, I. P. Kikot, L. I. Manevitch and A. V. Onufriev, *Phys. Rev. B*, 2011, **83**, 245406.
- 65 C. W. Hsu, M. Fyta, G. Lakatos, S. Melchionna and E. Kaxiras, *J. Chem. Phys.*, 2012, **137**, 105102.
- 66 T. E. Ouldridge, A. A. Louis and J. P. K. Doye, *J. Chem. Phys.*, 2011, **134**, 085101.
- 67 L. E. Edens, J. A. Brozik and D. J. Keller, *J. Phys. Chem. B*, 2012, **116**, 14735–14743.
- 68 M. Maciejczyk, A. Spasic, A. Liwo and H. A. Scheraga, *J. Comput. Chem.*, 2010, **31**, 1644–1655.

- 69 Y. He, M. Maciejczyk, S. Oldziej, H. A. Scheraga and A. Liwo, *Phys. Rev. Lett.*, 2013, **110**, 098101.
- 70 T. Cragoli, P. Derreumaux and S. Pasquali, *J. Phys. Chem. B*, 2013, **117**, 8047–8060.
- 71 A. A. Louis, *J. Phys.: Condens. Matter*, 2002, **14**, 9187–9206.
- 72 R. Everaers, S. Kumar and C. Simm, *Phys. Rev. E*, 2007, **75**, 041908.
- 73 D. Chandler and P. Varilly, *Proceeding of the International School of Physics "Enrico Fermi"*, IOS Press, Amsterdam, 2012, vol. 176, pp. 75–111.
- 74 C. R. Chandler, H. R. Drew, B. F. Luisi and A. A. Travers, *Understanding DNA: The molecule and how it works*, Elsevier Academic Press, 2004.
- 75 A.-M. Florescu and M. Joyeux, *J. Chem. Phys.*, 2011, **135**, 085105.
- 76 P. K. Maiti, T. A. Pascal, N. Vaidehi and W. A. Goddard III, *Nucl. Acids Res.*, 2004, **32**, 6047–6056.
- 77 P. K. Maiti, T. A. Pascal, N. Vaidehi, J. Heo and W. A. Goddard, *Biophys. J.*, 2006, **90**, 1463–1479.
- 78 J. M. Arbona, J.-P. Aimé and J. Elezgaray, *Phys. Rev. E*, 2012, **86**, 051912.
- 79 C. E. Castro, F. Kilchherr, E. L. Kim, D.-N. Shiao, T. Wauer, P. Wortmann, M. Bathe and H. Dietz, *Nat. Methods*, 2011, **8**, 221–229.
- 80 D.-N. Kim, F. Kilchherr, H. Dietz and M. Bathe, *Nucl. Acids Res.*, 2012, **40**, 2862–2868.
- 81 T. E. Ouldridge, *PhD thesis*, University of Oxford, 2011.
- 82 J. A. Holbrook, M. W. Capp, R. M. Saecker and M. T. Record, *Biochemistry*, 1999, **38**, 8409–8422.
- 83 P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. K. Doye and A. A. Louis, *J. Chem. Phys.*, 2012, **137**, 135101.
- 84 F. Romano, A. Hudson, J. P. K. Doye, T. E. Ouldridge and A. A. Louis, *J. Chem. Phys.*, 2012, **136**, 215102.
- 85 C. Matek, T. E. Ouldridge, A. Levy, J. P. K. Doye and A. A. Louis, *J. Phys. Chem. B*, 2012, **116**, 11616–11625.
- 86 F. Romano, D. Chakraborty, J. P. K. Doye, T. E. Ouldridge and A. A. Louis, *J. Chem. Phys.*, 2013, **138**, 085101.
- 87 N. Srinivas, T. E. Ouldridge, P. Šulc, J. M. Schaeffer, B. Yurke, A. A. Louis, J. P. K. Doye and E. Winfree, *Nucl. Acids Res.*, 2013, **41**, accepted.
- 88 T. E. Ouldridge, P. Šulc, F. Romano, J. P. K. Doye and A. A. Louis, *Nucl. Acids Res.*, 2013, **41**, in press; arXiv:1303.3370.
- 89 C. de Michele, L. Rovigatti, T. Bellini and F. Sciortino, *Soft Matter*, 2012, **8**, 8388–8398.
- 90 A. A. Kornyshev, D. J. Lee, S. Leikin and A. Wynveen, *Rev. Mod. Phys.*, 2007, **79**, 943–996.
- 91 F. Mocci and A. Laaksonen, *Soft Matter*, 2012, **8**, 9268–9284.
- 92 <http://dna.physics.ox.ac.uk>.
- 93 T. E. Ouldridge, *J. Chem. Phys.*, 2012, **137**, 144105.
- 94 I. Snook, *The Langevin and Generalised Langevin Approach to the Dynamics of Atomic, Polymeric and Colloidal Systems*, Elsevier, Amsterdam, 2007.
- 95 D. Frenkel and B. Smit, *Understanding Molecular Simulation*, Academic Press, San Diego, 2001.
- 96 R. D. Hills, L. Lu and G. A. Voth, *PLoS. Comp. Bio.*, 2010, **6**, e1000827.
- 97 K. Kikuchi, M. Yoshida, T. Maekawa and H. Watanabe, *Chem. Phys. Lett.*, 1991, **185**, 335–338.
- 98 L. Berthier and W. Kob, *J. Phys.: Condens. Matter*, 2007, **19**, 205130.
- 99 G. Tiana, L. Sutto and R. A. Broglia, *Physica A*, 2007, **380**, 241–249.
- 100 E. Sanz and D. Marenduzzo, *J. Chem. Phys.*, 2010, **132**, 194102.
- 101 S. Whitlam and P. Geissler, *J. Chem. Phys.*, 2007, **127**, 154101.
- 102 S. Whitlam, E. H. Feng, M. F. Hagan and P. Geissler, *Soft Matter*, 2009, **5**, 1251–1262.
- 103 A. Bhattacharyay and A. Troisi, *Chem. Phys. Lett.*, 2008, **458**, 210–213.
- 104 R. L. Davidchack, R. Handel and M. V. Tretyakov, *J. Chem. Phys.*, 2009, **130**, 234101.
- 105 J. Russo, P. Tartaglia and F. Sciortino, *J. Chem. Phys.*, 2009, **131**, 014504.
- 106 G. M. Torrie and J. P. Valleau, *J. Comp. Phys.*, 1977, **23**, 187.
- 107 D. J. Earl and M. W. Deem, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3910–3916.
- 108 T. E. Ouldridge, A. A. Louis and J. P. K. Doye, *J. Phys.: Condens. Matter*, 2010, **22**, 104102.
- 109 T. R. Prytkova, I. Eryazici, B. Stepp, S.-B. Nguyen and G. C. Schatz, *J. Phys. Chem. B*, 2010, **114**, 2627–2634.
- 110 M. J. Hoefert, E. J. Sambriski and J. J. de Pablo, *Soft Matter*, 2011, **7**, 560–566.
- 111 T. J. Schmitt and T. A. Knotts IV, *J. Chem. Phys.*, 2011, **134**, 205105.
- 112 D. H. D. Jong, L. V. Schäfer, A. H. D. Vries, S. J. Marrink, H. J. C. Berendsen and H. Grubmüller, *J. Comput. Chem.*, 2011, **32**, 1919–1928.
- 113 R. J. Allen, P. B. Warren and ten Wolde P. R., *Phys. Rev. Lett.*, 2005, **94**, 018104.
- 114 R. J. Allen, C. Valeriani and P. R. ten Wolde, *J. Phys.: Condens. Matter*, 2009, **21**, 463102.
- 115 C. Dellago, P. G. Bolhuis and P. L. Geissler, *Adv. Chem. Phys.*, 2002, **123**, 1–84.
- 116 T. S. van Erp and P. G. Bolhuis, *J. Comp. Phys.*, 2005, **205**, 157–181.
- 117 N. L. Goddard, G. Bonnet, O. Krichevsky, and A. Libchaber, *Phys. Rev. Lett.*, 2000, **85**, 2400–2403.
- 118 E. J. Sambriski, D. C. Schwartz and J. J. de Pablo, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 18125–18130.
- 119 T. J. Schmitt, J. B. Rogers and T. A. Knotts IV, *J. Chem. Phys.*, 2013, **138**, 035102.
- 120 N. R. Cozzarelli, G. J. Cost, M. Nöllmann, T. Viard and J. E. Stray, *Nat. Rev. Mol. Cell Biol.*, 2006, **53**, 580–588.
- 121 J. M. Fogg, G. L. Randall, B. M. Pettitt, D. W. L. Summners, S. A. Harris and L. Zechiedrich, *Quart. Rev. Biophys.*, 2012, **45**, 257–299.
- 122 J. D. Moroz and P. Nelson, *J. Chem. Phys.*, 1997, **94**, 14418–14422.
- 123 J. R. Wenner, M. C. Williams, I. Rouzina and V. A. Bloomfield, *Biophys. J.*, 2002, **82**, 3160–3169.
- 124 X. Zhang, H. Chen, H. Fu, P. S. Doyle and J. Yan, *Proc. Natl. Acad. Sci. USA*, 2012, **109**, 8103–8108.
- 125 M. C. Williams, I. Rouzina and V. A. Bloomfield, *Acc. Chem. Res.*, 2002, **35**, 159–166.
- 126 S. B. Smith, Y. Cui and C. Bustamante, *Science*, 1996, **271**, 795–799.
- 127 G. A. King, P. Gross, U. Bockelmann, M. Modesti, G. J. L. Wuite and E. J. G. Peterman, *Proc. Natl. Acad. Sci. USA*, 2013, **110**, 3859–3864.
- 128 X. Zhang, H. Chen, I. Rouzina, P. S. Doyle and J. Yan, *Proc. Natl. Acad. Sci. USA*, 2013, **110**, 3865–3870.
- 129 J. M. Hugué, C. V. Bizarro, N. Forns, S. B. Smith, C. Bustamante and F. Ritort, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 15431–15436.
- 130 W.-S. Chen, W.-H. Chen, Z. Chen, A. A. Gooding, K.-J. Lin and C.-H. Kiang, *Phys. Rev. Lett.*, 2010, **105**, 218104.
- 131 T. R. Strick, J.-F. Allemand, D. Bensimon, A. Bensimon and V. Croquette, *Science*, 1996, **271**, 1835–1837.
- 132 A. Vologodskii and M. D. Frank-Kamenetskii, *Nucleic Acids Res.*, 2013, **41**, 6785–6792.
- 133 D. Shore, J. Langowski and R. L. Baldwin, *Proc. Natl. Acad. Sci. USA*, 1981, **78**, 4833–4837.
- 134 T. E. Cloutier and J. Widom, *Mol. Cell*, 2004, **14**, 355–362.
- 135 J. Yan and J. F. Marko, *Phys. Rev. Lett.*, 2004, **93**, 108108.
- 136 D. A. Sivak and P. L. Geissler, *J. Chem. Phys.*, 2012, **136**, 045102.
- 137 H. Shroff, D. Sivak, J. J. Siegel, A. L. McEvoy, M. Siu, A. Spakowitz, P. L. Geissler and J. Liphardt, *Biophys. J.*, 2008, **94**, 2179–2186.
- 138 H. Qu, Y. Wang, C.-Y. Tseng and G. Zocchi, *Phys. Rev. X*, 2011, **1**, 021008.
- 139 D. J. Kauert, T. Kurth, T. Liedl and R. Seidel, *Nano Lett.*, 2011, **11**, 5558–5563.

-
- 140 E. Pfitzner, C. Wachauf, F. Kilchherr, B. Pelz, W. M. Shih, M. Rief and H. Dietz, *Angew. Chem. Int. Ed.*, 2013, **52**.
- 141 J. Nowakowski, P. J. Shim, C. D. Stout and G. F. Joyce, *J. Mol. Biol.*, 2000, **300**, 93–102.
- 142 D. M. J. Lilley, *Quart. Rev. Biophys.*, 2000, **33**, 109–159.
- 143 R. P. Goodman, I. A. T. Schaap, C. F. Tardin, C. M. Erben, R. M. Berry, C. F. Schmidt and A. J. Turberfield, *Nature*, 2005, **310**, 1661–1665.
- 144 R. M. Dirks, J. S. Bois, J. M. Schaeffer, E. Winfree and N. A. Pierce, *SIAM Rev.*, 2007, **49**, 65–88.
- 145 T.-J. Fu and N. C. Seeman, *Biochemistry*, 1993, **32**, 3211–3220.
- 146 E. Winfree, F. Liu, L. A. Wenzler and N. C. Seeman, *Nature*, 1998, **394**, 539–544.
- 147 P. W. K. Rothemund, A. Ekani-Nkodo, N. Papadakis, A. Kumar, D. K. Fygenson and E. Winfree, *J. Am. Chem. Soc.*, 2004, **126**, 16344–16353.
- 148 P. W. K. Rothemund, N. Papadakis and E. Winfree, *PLoS Biology*, 2004, **2**, e424.
- 149 S. R. and E. Winfree, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 15236–15241.
- 150 Y. He, M. Su, P. Fang, C. Zhang, A. E. Ribbe, W. Jiang and C. Mao, *Angew. Chem. Int. Ed.*, 2010, **49**, 748–751.
- 151 C. Zhang, M. Su, Y. He, X. Zhao, P.-A. Fang, A. E. Ribbe, W. Jiang and C. Mao, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 10665–10669.
- 152 Y. He, Y. Chen, H. Liu, A. E. Ribbe and C. Mao, *J. Am. Chem. Soc.*, 2005, **127**, 12202–12203.
- 153 H. Yan, S. H. Park, G. Finkelstein, J. H. Reif and T. H. LaBean, *Science*, 2003, **301**, 1882–1884.
- 154 Y. He, Y. Tian, A. E. Ribbe and C. Mao, *J. Am. Chem. Soc.*, 2006, **128**, 15978–15979.
- 155 Y. He, T. Ye, C. Zhang, A. E. Ribbe, W. Jiang and C. Mao, *Nature*, 2008, **452**, 198–201.
- 156 P. W. K. Rothemund, *Nature*, 2006, **440**, 297–302.
- 157 S. M. Douglas, H. Dietz, T. Liedl, B. Högbert, F. Graf and W. M. Shih, *Nature*, 2009, **459**, 414–418.
- 158 H. Dietz, S. M. Douglas and W. M. Shih, *Science*, 2009, **325**, 725730.
- 159 D. Han, S. Pal, J. Nangreave, Z. Deng, Y. Liu and H. Yan, *Science*, 2011, **332**, 342–346.
- 160 X.-C. Bai, T. G. Martin, S. H. W. Scheres and H. Dietz, *Proc. Natl. Acad. Sci. USA*, 2012, **109**, 20012–20017.
- 161 S. M. Douglas, A. H. Marblestone, S. Teerapittayanon, A. Vazquez, G. M. Church and W. M. Shih, *Nucl. Acids Res.*, 2009, **37**, 5001–5006.
- 162 J.-P. Sobczak, T. G. Martin, G. Gerling and H. Dietz, *Science*, 2012, **338**, 1458–61.
- 163 Y. Yang, D. Han, J. Nangreave, Y. Liu and H. Yan, *ACS Nano*, 2012, **6**, 8209–8215.
- 164 J. S. Bois, S. Venkataraman, H. M. T. Choi, A. J. Spakowitz, Z.-G. Wang and N. A. Pierce, *Nucleic Acids Res.*, 2005, **33**, 4090–4095.
- 165 S. J. Green, D. Lubrich and A. J. Turberfield, *Biophys. J.*, 2006, **91**, 2966–2975.
- 166 J. Bath, S. J. Green and A. J. Turberfield, *Angew. Chem. Int. Ed.*, 2005, **117**, 4432–4435.
- 167 S. F. Wickham, J. Bath, Y. Katsuda, M. Endo, K. Hidaka, H. Sugiyama and A. J. Turberfield, *Nature Nanotechnology*, 2012, **7**, 169–173.
- 168 J. Bath, S. J. Green, K. E. Allen and A. J. Turberfield, *Small*, 2009, **5**, 1513–1516.
- 169 S. J. Green, J. Bath and A. J. Turberfield, *Phys. Rev. Lett.*, 2008, **101**, 238101.