AIP | The Journal of Chemical Physics

Introducing improved structural properties and salt dependence into a coarse-grained model of DNA

Benedict E. K. Snodin, Ferdinando Randisi, Majid Mosayebi, Petr Šulc, John S. Schreck, Flavio Romano, Thomas E. Ouldridge, Roman Tsukanov, Eyal Nir, Ard A. Louis, and Jonathan P. K. Doye

Citation: The Journal of Chemical Physics **142**, 234901 (2015); doi: 10.1063/1.4921957
View online: http://dx.doi.org/10.1063/1.4921957
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/142/23?ver=pdfcov
Published by the AIP Publishing

Articles you may be interested in

Equilibration of complexes of DNA and H-NS proteins on charged surfaces: A coarse-grained model point of view
J. Chem. Phys. **141**, 115102 (2014); 10.1063/1.4895819

A nucleotide-level coarse-grained model of RNA
J. Chem. Phys. **140**, 235102 (2014); 10.1063/1.4881424

Sequence-dependent thermodynamics of a coarse-grained DNA model
J. Chem. Phys. **137**, 135101 (2012); 10.1063/1.4754132

Influence of mobile DNA-protein-DNA bridges on DNA configurations: Coarse-grained Monte-Carlo simulations
J. Chem. Phys. **135**, 125104 (2011); 10.1063/1.3636383

Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model
J. Chem. Phys. **134**, 085101 (2011); 10.1063/1.3552946

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22

THE JOURNAL OF CHEMICAL PHYSICS **142**, 234901 (2015)

Introducing improved structural properties and salt dependence into a coarse-grained model of DNA

Benedict E. K. Snodin,[1,a)] Ferdinando Randisi,[2,3] Majid Mosayebi,[1] Petr Šulc,[4] John S. Schreck,[1] Flavio Romano,[1] Thomas E. Ouldridge,[5] Roman Tsukanov,[6] Eyal Nir,[6] Ard A. Louis,[3] and Jonathan P. K. Doye[1,b)]

[1]*Physical and Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QZ, United Kingdom*
[2]*Life Sciences Interface Doctoral Training Center, South Parks Road, Oxford OX1 3QU, United Kingdom*
[3]*Rudolf Peierls Centre for Theoretical Physics, 1 Keble Road, Oxford OX1 3NP, United Kingdom*
[4]*Center for Studies in Physics and Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA*
[5]*Department of Mathematics, Imperial College, 180 Queen's Gate, London SW7 2AZ, United Kingdom*
[6]*Department of Chemistry and the Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev, Beer Sheva, Israel*

(Received 2 April 2015; accepted 8 May 2015; published online 15 June 2015)

We introduce an extended version of oxDNA, a coarse-grained model of deoxyribonucleic acid (DNA) designed to capture the thermodynamic, structural, and mechanical properties of single- and double-stranded DNA. By including explicit major and minor grooves and by slightly modifying the coaxial stacking and backbone-backbone interactions, we improve the ability of the model to treat large (kilobase-pair) structures, such as DNA origami, which are sensitive to these geometric features. Further, we extend the model, which was previously parameterised to just one salt concentration ($[Na^+] = 0.5$M), so that it can be used for a range of salt concentrations including those corresponding to physiological conditions. Finally, we use new experimental data to parameterise the oxDNA potential so that consecutive adenine bases stack with a different strength to consecutive thymine bases, a feature which allows a more accurate treatment of systems where the flexibility of single-stranded regions is important. We illustrate the new possibilities opened up by the updated model, oxDNA2, by presenting results from simulations of the structure of large DNA objects and by using the model to investigate some salt-dependent properties of DNA. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4921957]

I. INTRODUCTION

Deoxyribonucleic acid (DNA) performs the crucial function of storing genetic information in living organisms. It is made up of repeating units called nucleotides, each of which consists of a sugar and phosphate backbone plus a base (either Adenine (A), Guanine (G), Thymine (T), or Cytosine (C)) attached to the sugar. Watson-Crick base pairing, A with T and G with C, along with planar stacking interactions between bases and the constraints of the backbone, leads to the formation of the well-known double-helical structure of DNA.

The Watson-Crick complementarity of DNA also permits the rational design of DNA objects for which the intended structure is the global free-energy minimum, a property which has been exploited to create a wide variety of 2D and 3D nanostructures.[1–4] Further, these DNA objects can be functionalised,[5–9] with potential applications ranging from nanomedicine to nanoelectronics.[10]

Theoretical and computational approaches to modelling DNA have been widely exploited to probe the behaviour of the molecule in both a biological and a nanotechnological context. At the finest level of detail, quantum chemistry calculations have been used to study the interactions between nucleotides,[11–13] although the high computational cost of such an approach limits these methods to interactions between nearest-neighbour base pairs in vacuum. Classical all-atom approaches, where every atom of DNA and the surrounding solvent is modelled as a point particle with effective interactions, have been widely employed to study small DNA motifs[14,15] and have recently been applied to larger DNA systems.[16–18] However, simulating rare-event processes such as the breaking or formation of base pairs remains a challenge with these models, with μs time scales being the limit of what is currently accessible.[18] At the other end of the scale, theoretical approaches have been developed to understand certain large-scale properties of DNA. These include the wormlike-chain model, which treats DNA as a continuously flexible polymer.[19] While such models can provide useful insights into the physical properties of DNA, they are not detailed enough, by design, to address processes such as duplex formation.

The middle ground between analytical and all-atom approaches is occupied by coarse-grained DNA models. Such models integrate out many of the degrees of freedom of the DNA nucleotide and often neglect the solvent molecules; these approximations inevitably imply a compromise between

a)Electronic mail: benedict.snodin@chem.ox.ac.uk
b)Electronic mail: jonathan.doye@chem.ox.ac.uk

 © 2015 AIP Publishing LLC

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22

accuracy, generality, and computational efficiency, so that care must be taken in applying these models to a given problem. However, the simplified picture presented by such an approach can be a strength, as in addition to greatly increasing the time scale and number of nucleotides that can be studied, it can allow one to understand the generic physics governing the system more easily.

Many coarse-grained DNA models have been developed in recent years;[20–29] in this work, we improve the model due to Ouldridge *et al.*, oxDNA.[30,31] This nucleotide-level model is designed with a heuristic, "top-down" approach, with a focus on reproducing well-known properties of DNA (such as the helical structure of the B-DNA duplex) and experimental results (such as duplex melting temperatures), rather than, for example, building the model up by integrating out details from an all-atom representation. In the oxDNA model, each nucleotide is a rigid body with three interaction sites that have mutual, highly anisotropic interactions. This treatment is sufficient to obtain good agreement with experimental data on the structural, mechanical, and especially the thermodynamic properties of single- and double-stranded DNA. Consequently, the model has provided key insights into many different processes relevant to DNA nanotechnology[32–40] and biophysics[41–45] and importantly has also been shown to provide direct agreement with experimentally measured properties on a range of systems including DNA overstretching,[45] a two-footed DNA walker,[35] and toehold-mediated strand displacement.[39,46]

Despite these achievements, there are some areas where oxDNA can be improved. The model was parameterised to $[Na^+] = 0.5$M, a high salt concentration similar to that used for many applications in DNA nanotechnology. However, the ability to study DNA behaviour as a function of salt would allow quantitative comparison with a greatly expanded set of experiments, and, in particular, if we wish to apply oxDNA to biological systems, we would like to work at physiological salt ($[Na^+] \approx 0.15$M). At the same time, the wealth of experimental data for the thermodynamic and mechanical properties of DNA as a function of salt concentration[47–50] makes fitting the salt-dependent properties of an extended model possible.

While the detailed effect of salt on DNA electrostatics can be highly complex,[51,52] here we use a simple Debye-Hückel interaction term, as first implemented for oxDNA by Wang and Pettitt in Ref. 42, to model how salt screens repulsive interactions. This more coarse-grained description is commensurate with the level of approximation used more generally in the oxDNA framework. We carefully parameterise the model to an extended set of experimental data for melting temperatures and persistence lengths, including the behaviour of single strands.

A second area that merits attention is the performance of the model in simulating the structure of large (kilobase-pair) DNA objects. The model value of the B-DNA pitch did not come under too much scrutiny in the original parameterisation as there is some disagreement in the literature about the precise value of the average pitch.[50,53–57] However, improved experimental techniques in fabrication and imaging of large DNA objects[2,58,59] have presented an opportunity to finely tune this value, because small adjustments to the duplex pitch can result in significant changes to the global twist of a large-scale DNA nanostructure. In addition, simulating these large-scale structures has illustrated the potential importance of nicks and junctions for the effective duplex pitch, so that improving the model's description of these effects has become a priority. Finally, the original model duplex had grooves with equal widths, whereas B-DNA is known to have a larger major groove and a smaller minor groove. This implies that the positions of the backbone sites in the model, which are directly related to the groove widths, could be more realistic. This detail could be relevant, for example, in origami structure, where the precise backbone positions are known to be important for junction placement.[2]

The oxDNA model was previously given sequence-dependent hydrogen-bonding and stacking strengths[60] by fitting to the duplex thermodynamics of the SantaLucia model.[49] As the SantaLucia model gives results at the base-pair step level, one can only extract the average stacking strength for the two stacking interactions present in a given base-pair step. In particular, this means that in oxDNA, the AA and TT stacking interactions are the same, whereas it is well known that the AA stacking interaction is significantly stronger than the TT interaction, an important property for DNA nanotechnology, for example, where poly-T single-stranded regions are often used as flexible linkers.[61,62] To remedy this, we use new experimental data to reparameterise the AA and TT stacking interactions in the model.

In Secs. II–VIII, after briefly introducing the original oxDNA model, we consider each change to the model in turn, and then we highlight some important aspects of the behaviour of the new model.

II. THE ORIGINAL oxDNA MODEL

In the oxDNA model introduced by Ouldridge *et al.*,[30] a strand of DNA is represented by a chain of rigid bodies, with each rigid body representing a nucleotide. The coarse-grained, pairwise potential for the model can be written as a sum of interaction terms

$$V = \sum_{\text{nearest neighbours}} (V_{\text{backbone}} + V_{\text{stack}} + V'_{\text{exc}}) + \sum_{\text{other pairs}} (V_{\text{HB}} + V_{\text{cross stack}} + V_{\text{exc}} + V_{\text{coax stack}}). \quad (1)$$

The nearest neighbours (adjacent nucleotides on a DNA strand) interact via V_{backbone}, V_{stack}, and V'_{exc}, which represent the connectivity between neighbouring backbones, the favourable stacking interactions between neighbouring bases, and excluded volume terms, respectively. All other nucleotide pairs interact with V_{HB}, $V_{\text{cross stack}}$, V_{exc}, and $V_{\text{coax stack}}$, corresponding to hydrogen bonding between complementary bases, cross-stacking, excluded volume, and coaxial stacking between non-nearest neighbours, respectively. We now highlight aspects of the original model which are relevant to the improvements made in this paper.

First, in its original formulation, the model was parameterised for a sodium ion concentration of 0.5M, chosen to reflect the high salt concentrations commonly used in DNA nanotechnology. The electrostatic interactions between the negatively charged phosphate groups on the DNA backbone

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22

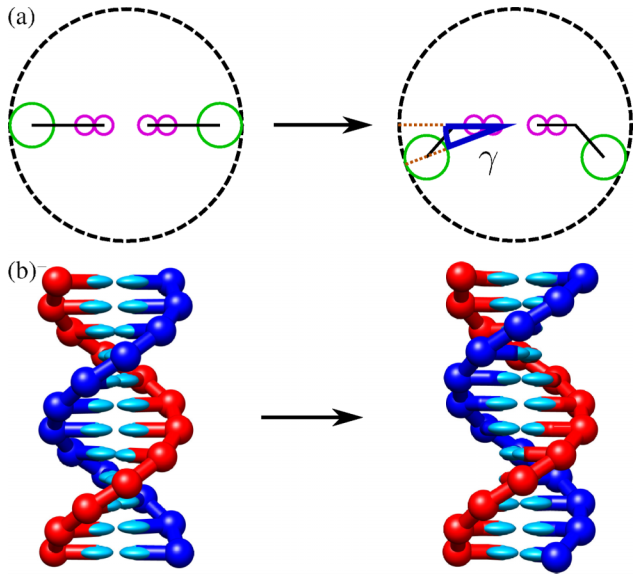

FIG. 1. Schematics contrasting the original oxDNA model (left), with equal groove widths, with oxDNA2 (right), which has differentiated major and minor grooves. (a) A cross section of a duplex with one base pair displayed. The large dashed circle shows the helix radius, and each nucleotide is represented by three circles joined by a line; the large solid circles represent the backbone sites, while the small solid circles represent the stacking (closer to the backbone) and hydrogen-bonding (at the end of the nucleotide) sites. For oxDNA2, a value of 20° was chosen for the angle γ. (b) A representation of a DNA duplex for each model.

were incorporated into the backbone site's excluded volume, an approximation which can be justified by the very short Debye screening length at that relatively high ion concentration.

Second, the original oxDNA model[30] represented each nucleotide as a linear rigid body (Fig. 1). The optimal configuration for base-pairing occurs when the two nucleotides point directly at each other. As a consequence, the DNA double helix was symmetric, with the two grooves having equal widths.

Third, in the original oxDNA model introduced in Ref. 30, all four types of base were treated equally except that only A-T and G-C base pairs could be formed. Later, Šulc *et al.*[60] introduced sequence-dependent thermodynamics into the model by making the strengths of the hydrogen-bonding and stacking terms depend on the identities of the interacting nucleotides. The nearest-neighbour DNA model of SantaLucia,[49] to which oxDNA was parameterised, does not resolve the difference between AA and TT stacking as it works on the base-pair step level—therefore AA and TT stacking strengths were set to be the same in Ref. 60.

In this work, we mostly work from the original, sequence-averaged parameterisation of the model rather than the sequence-dependent one, as it is more efficient to fit the thermodynamic parameters to sequence-averaged duplex melting temperatures as given by the SantaLucia model. The exception is the parameterisation of the AA and TT stacking strengths, which did use the sequence-dependent parameters from Ref. 60 as a starting point, to allow the best possible comparison between the model and the experimental results that were used for the fitting. After the parameters for oxDNA2 had been obtained, including new values for the sequence-averaged hydrogen bonding and stacking strengths, we then rescaled the sequence-dependent interaction strengths from Ref. 60 accordingly for use with the new model.

III. INTRODUCING DIFFERENT WIDTHS FOR MAJOR AND MINOR DNA GROOVES

B-DNA in the original oxDNA model has equal groove widths, while in reality DNA has a larger major groove and a smaller minor groove. Having realistic widths for the major and minor grooves is equivalent to having appropriately positioned backbone sites in the model, an important feature for the physical properties of many DNA motifs. For example, in DNA origami, antiparallel double helices are joined by crossovers, for which the position of the backbone has been shown to be crucial for origami structure.[2,63] Another example is anisotropic duplex bending: the duplex can be expected to bend more easily into the major groove than into the minor, if the groove widths are unequal.

The oxDNA nucleotide is composed of three interaction sites: the hydrogen-bonding, stacking, and backbone sites. We introduce different groove widths by changing the position of the backbone site while keeping the duplex radius unchanged (Fig. 1), such that, rather than lying on a straight line, the three interaction sites lie in a plane. The new nucleotide shape introduces an additional parameter into the model, the angle γ between the line from the duplex centre to the backbone site and the line from the duplex centre to the stacking site (Fig. 1(a)). Given the coarse graining of the 18 atoms of the sugar-phosphate DNA backbone into a single interaction site, there is no definitive choice for the precise position of the backbone site and thus the value of the model parameter γ (Fig. 1(a)). We set $\gamma = 20°$, a value which maps onto a full-atom representation of a DNA duplex well by visual inspection, although values of γ between 15° and 25° would give an equally satisfying visual match.

The backbone site is moved such that the duplex radius is unchanged, and we note that the modification has a negligible added computational cost when simulating the model. However, the thermodynamic and mechanical properties are slightly affected. For the thermodynamics, we found a change of 1-2 K in the duplex melting temperatures, and we modified the hydrogen-bonding and stacking strengths using the histogram reweighting method described in Sec. II C 2 of Ref. 64, so that the agreement with experimental melting temperatures was as good as for the original model. The mechanical properties of DNA are less well constrained experimentally and so were not refitted. The mechanical properties for the new model can be found in Sec. VIII.

One illustration of the importance of the groove widths in oxDNA for the structural properties of DNA assemblies is provided by systems of 3-arm star tiles that are designed to form triangular prismatic polyhedra. We find that modifying the groove widths qualitatively changes the structure of trimers of these tiles (Fig. 2). Specifically, the body of the trimer defines a plane with two distinct faces. Zhang *et al.*[65] found that one of two possible isomers of the polyhedron preferentially formed, implying that the free arms of the trimer systematically pointed in the direction of one of these two faces. We find a consistent result when the groove widths specified by oxDNA2 are used.

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22

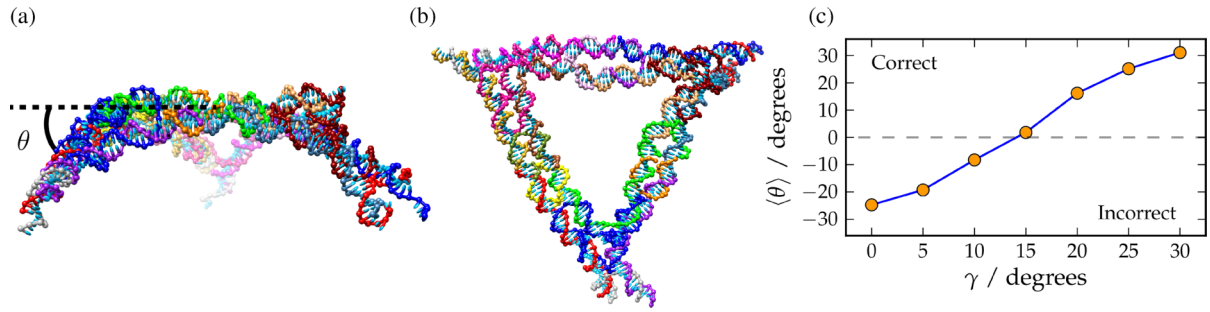

FIG. 2. In oxDNA, the structure of the DNA trimer due to Zhang *et al.*[65] is found to qualitatively change as a function of duplex groove angle γ. A representative trimer configuration with $\gamma = 20°$ is shown in (a) from the side and with θ, the angle between each arm and the plane of the main, triangular trimer section, displayed and in (b) from the top. (c) shows the average value of θ as a function of γ, with $\gamma = 0°$ corresponding to equal groove widths as for the original oxDNA model and $\gamma = 20°$ being chosen for oxDNA2.

When equal-sized grooves are used (as in the original oxDNA model), the trimer arms point in the opposite direction.

IV. EFFECTIVE ELECTROSTATIC INTERACTIONS

One major improvement presented in this paper is the introduction of a salt-dependent interaction term in the oxDNA model. Since the original goal of the oxDNA model was to simulate nanotechnology experiments, the thermodynamic and structural parameterisations were carried out at high salt. The very short electrostatic screening length at these conditions allows one to incorporate the electrostatics into a soft excluded volume, somewhat circumventing the necessity of a proper treatment. The original parameterisation of the thermodynamics was carried out at 0.5M [Na^+], a high enough value that further increasing it does not significantly change the physics at our level of coarse-graining.

The problem of treating electrostatics properly for DNA in solution is a complicated one. Perhaps, the most evident issue is that the typical dimensions of nucleotides are comparable to the Debye length of the solution, rendering a mean-field treatment hard to justify. Also, in some cases, the presence of salt ions affects the local structure of nucleic acids, by stabilizing some arrangements or destabilizing others. Ion condensation may also lead to stronger screening of the electrostatic interactions than otherwise expected—this has been incorporated into coarse-grained DNA models through partial effective charges.[66,67] Thus, in principle, many non-trivial effects must be taken into account when modelling the electrostatic interactions, and the debate on the best way to do so implicitly is still unresolved.[51,52]

Here, we choose a very simple treatment based on the Debye-Hückel model for screened electrostatics. This approach has been used previously for other coarse-grained models[25,66] and was first introduced into oxDNA by Wang and Pettitt.[42] We note that this treatment is consistent with the coarse-grained nature of the model, and we use the same top-down strategy that was used in the original parameterisation to design the effective electrostatic interactions: Our goal is to introduce a term in the potential that will reproduce the thermodynamic and mechanical effects of salt concentration on DNA and thus should be regarded as an effective interaction rather than an attempt to rigorously model the local effects of charges. This should be kept in mind when interpreting the results obtained with oxDNA, especially at low salt concentration.

Since the modelling of electrostatics is rather crude, we restrict our parameterisation to salt concentrations of 0.1M of monovalent salt or greater. This restriction is also due to the fact that we parameterise our thermodynamics to the model of SantaLucia,[49] which was fitted in a similar salt regime. Importantly, physiological conditions fall within this range, which will allow quantitative comparison between simulations of DNA systems with our model and experiments at physiological conditions.

The detailed Debye-Hückel form, which is added to the non-bonded interactions in the potential of Eq. (1), takes the following form:

$$V_{\mathrm{DH}}(T,I) = \sum_{ij} \frac{(q_{\mathrm{eff}}e)^2}{4\pi\epsilon_0\epsilon_{\mathrm{r}}} \frac{\exp\left\{-r_{ij}^{\text{b-b}}/\lambda_{\mathrm{DH}}(T,I)\right\}}{r_{ij}^{\text{b-b}}}, \tag{2}$$

where q_{eff} is the effective charge situated at the backbone site of each of the nucleotides, $r_{ij}^{\text{b-b}}$ is the distance between the backbone sites of nucleotides i and j, ϵ_0 is the permittivity of the vacuum, ϵ_r is the relative permittivity of water, e is the elementary charge, and q_{eff} is a dimensionless effective charge. In principle, ϵ_r depends on r_{ij}[68] and weakly depends on temperature and salt concentration. However, for oxDNA2, we set ϵ_r to be a constant value, in keeping with the coarse-grained approach taken for the rest of the model. In particular, we choose $\epsilon_r = 80$, the standard value for water. In Eq. (2), we have stressed that the interaction depends on the temperature T and on the (monovalent) salt concentration I through the Debye length $\lambda_{\mathrm{DH}}(T,I)$,

$$\lambda_{\mathrm{DH}}(T,I) = \sqrt{\frac{\epsilon_0\epsilon_{\mathrm{r}}k_{\mathrm{B}}T}{2N_{\mathrm{A}}e^2I}}, \tag{3}$$

where N_{A} is Avogadro's number and k_{B} is Boltzmann's constant.

To ensure the computational efficiency of simulating the model, we set the interaction to zero at a finite distance, and to allow simulation with molecular dynamics (MD), we introduce a quadratic smoothing potential so that the interaction goes to zero smoothly. The quadratic smoothing, the details of which are reported in Sec. I A of the supplementary material,[69] is introduced after a cutoff distance r_{smooth}, which we choose to be $3\lambda_{\mathrm{DH}}$. This cutoff allows us to use all the standard techniques to improve the simulation efficiency via the use of Verlet lists and/or cells.[70] We have checked that introducing

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22

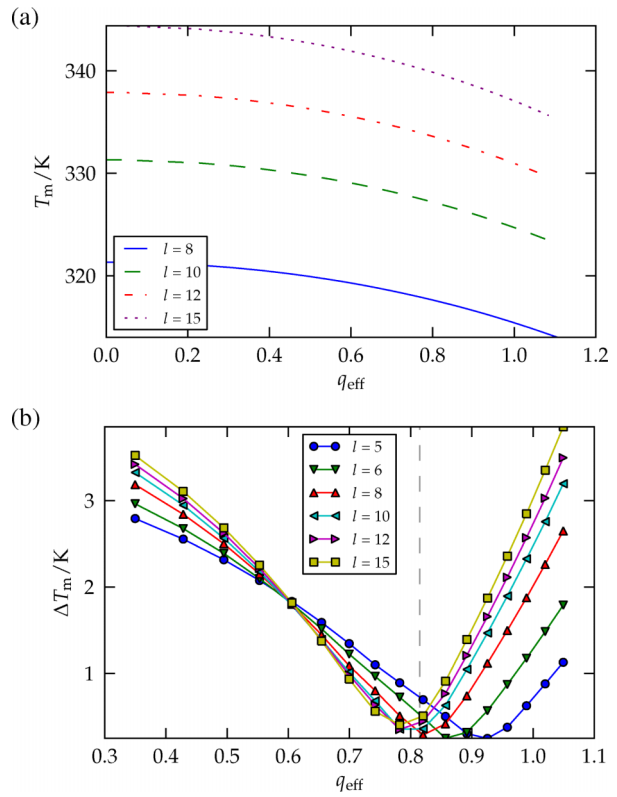

FIG. 3. (a) Melting temperature of duplex DNA of different lengths l at $[Na^+] = 0.2$M as a function of the effective charge q_{eff} obtained with thermodynamic integration. (b) Average melting temperature difference between our model and the SantaLucia model[49] for different values of l as a function of q_{eff}. Each point corresponds to the average over $[Na^+] = 0.1$, 0.2, 0.3, 0.4, and 0.5M of the magnitude of the difference in melting temperature between our model and the SantaLucia model. The plot suggests an optimal value for q_{eff} of 0.815, indicated by the vertical dashed line.

our chosen cutoff, $r_{smooth} = 3\lambda_{DH}$, has a negligible effect on the duplex thermodynamics results used to parameterise the interaction (Fig. 3(a)).

Our representation of DNA uses a single rigid body per nucleotide, and the best choice of where to put the charge is not obvious. All the atoms of the sugar and phosphate groups of the backbone are represented in a single interaction site, and it is thus natural to put the charge, which in real DNA is located on the phosphate, on that interaction site. Importantly, the backbone site of a nucleotide is placed almost in between the phosphate of that nucleotide and the phosphate of the neighbouring one, which could potentially lead to some unphysical effects. Also, we stress that having a charge at each backbone site means that the DNA has as many charges as nucleotides, which is not always true in real systems: very often, the terminal phosphate at the 3′ end is cut off, removing a charge. The absence of this charge can cause measurable effects on the thermodynamics, and indeed the SantaLucia model[49] requires the presence or absence of the terminal phosphates as an input parameter. In keeping with our coarse-graining approach, we put a half effective charge on each of the terminal nucleotides to incorporate the fact that each charge should be halfway in between our backbone sites, emulating a system with the terminal charge removed, and parameterise to the SantaLucia model in a way consistent with this approach.

The parameter that we have tuned to reproduce the thermodynamics predicted by the model of SantaLucia is the effective charge q_{eff}. To do this, we used thermodynamic integration[71] to compute the melting temperatures of duplexes of length 5, 6, 7, 8, 10, 12, and 15 as a function of q_{eff} at several salt concentrations, and chose the value that best reproduced the melting temperatures predicted by SantaLucia's model.

The melting temperature of a duplex is defined as the temperature at which half of the strands in a stoichiometric bulk solution are in the duplex state. We cannot simulate the bulk system with oxDNA. However, at the bulk melting temperature for a given concentration of strands, the bound and unbound free energies obtained for a system of two strands at the same concentration, F_b and F_{ub}, are related by[72]

$$F_b(T_m, q_{eff}) = F_{ub}(T_m, q_{eff}) - k_B T_m \ln(2), \tag{4}$$

where we have made explicit the dependence of F on the effective charge q_{eff}. The constant on the right hand side of Eq. (4) accounts for concentration fluctuations that are present in bulk but not for two strands in a periodic box.[72] The equation is exact in the dilute limit, which is indeed where the thermodynamics of duplex formation are usually studied, and is the relevant limit for oxDNA.

We can take advantage of the relation of the free energies given by Eq. (4) for the thermodynamic integration. For small changes in T and q_{eff}, it is possible to write

$$\begin{aligned} &F(T_m + dT, q_{eff} + dq_{eff}) \\ &\quad = F(T_m, q_{eff}) + \frac{\partial F}{\partial T} dT + \frac{\partial F}{\partial q_{eff}} dq_{eff} \\ &\quad = F(T_m, q_{eff}) + \left(\frac{F - \langle V \rangle}{T_m} + \left\langle \frac{\partial V}{\partial T} \right\rangle \right) dT + 2 \frac{\langle V_{DH} \rangle}{q_{eff}} dq_{eff}. \end{aligned} \tag{5}$$

The last step includes the unusual term $\langle \partial V/\partial T \rangle$. This is because our potential, like many coarse-grained models, depends explicitly on T through V_{stack} and V_{DH}.

To compute the change in T_m introduced by a small change in q_{eff}, one can impose the condition in Eq. (4) at the new T_m

$$\begin{aligned} F_b(T_m + dT, q_{eff} + dq_{eff}) = F_{ub}(T_m + dT, q_{eff} + dq_{eff}) \\ - k_B (T_m + dT) \ln(2), \end{aligned} \tag{6}$$

and, by using Eqs. (4)–(6), one obtains

$$\frac{dT_m}{dq_{eff}} = \frac{(2/q_{eff})(\langle V_{DH} \rangle_{ub} - \langle V_{DH} \rangle_b)}{(\langle V_{ub} \rangle - \langle V_b \rangle)/T_m - (\langle \partial V_{ub}/\partial T \rangle - \langle \partial V_b/\partial T \rangle)}. \tag{7}$$

Equation (7) is a differential equation that allows us to follow the change in melting temperature as q_{eff} is changed. We note that Eq. (7) is an extended Clausius-Clapeyron relation,[71] and the quantities on the right-hand side are readily accessible with separate simulations of the bound and unbound states. Since we are dealing with small systems, all these simulations are very quick and it is thus easy to achieve very accurate results. As a starting point for the thermodynamic integration, we use the melting temperatures from SantaLucia, which our original model (equivalent to the current model at $q_{eff} = 0$) reproduces within tenths of a Kelvin. Since both V_{stack} and V_{DH} only weakly depend on temperature, we used Eq. (7) assuming $\langle \partial V_{ub}/\partial T \rangle - \langle \partial V_b/\partial T \rangle = 0$ to obtain an optimal value for q_{eff}

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22

and checked *a posteriori* the results of thermodynamic integration with melting simulations. We have found that the melting temperature differences computed with Eq. (7) are accurate to within 0.2 K.

For a range of salt concentrations and duplex lengths, we performed thermodynamic integration to find the melting temperature as a function of q_{eff}. Some of the results are depicted in Fig. 3. The melting temperature of the duplex decreases with increasing effective charge q_{eff}, but not dramatically so (Fig. 3(a)). This is because the increased interchain repulsion in the duplex state is partially balanced by the lower entropy of the single-stranded state (due to it adopting a slightly more extended state to reduce intrachain repulsion). The data in Fig. 3(b) show that the q_{eff} at which the difference in melting temperature between oxDNA and SantaLucia is minimised depends on the duplex length and lies in the range $0.75 < q_{\mathrm{eff}} < 0.95$. The best overall predictions are obtained for an effective charge $q_{\mathrm{eff}} = 0.815$. It is reassuring that this value does not deviate significantly from 1 (corresponding to the value given by Debye-Hückel theory), and that the best value for q_{eff} varies little with duplex length. In addition, it is not uncommon to use a value of $q_{\mathrm{eff}} < 1$ for coarse-grained DNA models.[66,67] One argument is that ion condensation, which is known to occur for DNA, will screen the phosphate charges more strongly than expected from Debye-Hückel theory. This will lead to a lower effective charge when fitting a model using a Debye-Hückel treatment to experimental results, although such arguments should be applied with caution to a crude mean-field approach such as this one.

We note that introducing this explicit electrostatic term in our model potential will raise the computational expense of each simulation step compared to the original oxDNA model, as the electrostatic term will generally result in oxDNA2 having a larger interaction range than in the original model. This effect increases at lower salt, as the Debye-Hückel term becomes more long-ranged. For example, we find that simulating a 10-bp duplex with MD for a given number of steps takes 1.4× as long with oxDNA2 at $[\mathrm{Na}^+] = 0.5\mathrm{M}$ and 1.7× as long with oxDNA2 at $[\mathrm{Na}^+] = 0.1\mathrm{M}$ as it did for the original oxDNA.

V. IMPROVING STRUCTURE PREDICTION FOR LARGE-SCALE DNA OBJECTS

The model was originally parameterised for small single- and double-stranded DNA structures. If we wish to study large-scale structures, we need to ensure that we can reproduce the existing experimental data for these larger constructs. A good test case is provided by the work of Dietz *et al.*,[2] who measured the global twist of three different DNA origami structures, described as 10-by-6 helix bundles (Fig. 4). We denote the three origami structures as L-, N-, and R-type. In the experiment, they were designed to impose a pitch of 10, 10.5, and 11 base pairs per turn (bp/turn) on the constituent DNA double helices and these different designs exhibited left-handed, no, and right-handed global twist, respectively, when multimerised to form ribbons and visualised with transmission electron microscopy. One might think that this result implies that DNA has a natural pitch of 10.5 bp/turn, since the design with that inherent periodicity did not result in a globally twisted system. However, it is not that straightforward.

Unsurprisingly, when simulated with the original version of oxDNA, the N-type origami still showed a significant right-handed global twist, chiefly because the duplex pitch for the model was 10.36 bp/turn. However, even when simulated with a version of oxDNA modified so that the model duplex pitch was 10.5 bp/turn, the N-type origami displayed a right-handed twist. One reason for this is the following: in oxDNA, the helical twist across nicks and junctions is larger by about 3.9° and 2.5°, respectively, than for a normal duplex step (we call this an overtwist), so that fractionally fewer base pairs are required for a helix turn than would otherwise be expected. Although these differences are small, they add up constructively to create a global twist on a structure that would otherwise have no such twist. For a nicked duplex in oxDNA, the overtwist occurs because, opposite the nick, the non-nicked strand prefers a larger twist than for duplex DNA (just as stacked single strands do in the model), and the nicked backbones lack the finitely extensible nonlinear elastic (FENE) spring that would usually oppose this tendency. A similar argument applies for strands with junctions.

Although there is no direct evidence available to show whether this overtwist is physically realistic, there are multiple lines of experimental evidence, which suggest that the pitch of duplex DNA is close to 10.5 bp/turn.[50,53] Given the evidence from Dietz *et al.*[2] that the effective pitch in DNA origami structures is also close to 10.5 bp/turn, we decided to reduce the overtwisting in oxDNA as much as is possible. To achieve this, we modify the coaxial stacking term of the potential, $V_{\mathrm{coax\ stack}}$, so that the overtwist is 0° for a junction and 1.3° for a nick (see Sec. I C of the supplementary material[69] for details of the changes to the potential). We have verified that this change has very little effect on the other features of the model.

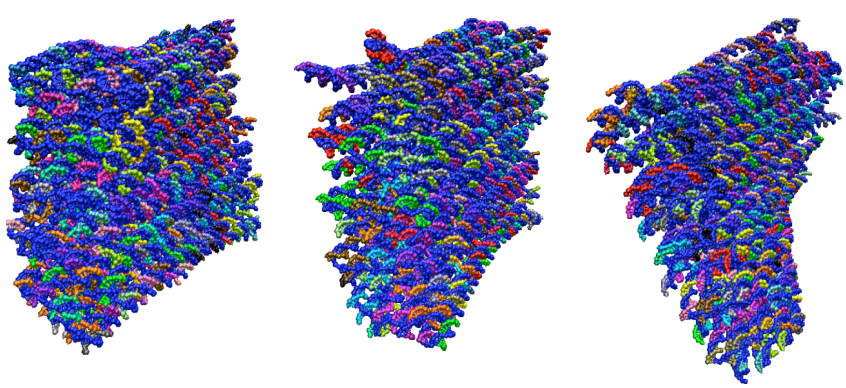

FIG. 4. Simulation snapshots for the L- (left), N- (middle), and R- (right) type helix bundles, each composed of roughly 15 000 nucleotides. The designs are taken from Dietz *et al.*[2] In experiment and in oxDNA2, the N-type helix bundle exhibits close to zero global twist.

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22

TABLE I. Global twist in the DNA helix bundles of Dietz *et al.*[2] for simulations using oxDNA2 and experiment. The oxDNA2 simulations were run with $[Na^+] = 0.5$M for 1.8×10^9 MD steps for each design.

	Global twist/deg	
Helix bundle type	oxDNA2	Experiment
N	0.2	0
L	−22.2	−31 ± 5
R	26.9	26 ± 5

Even when the overtwist at nicks and junctions has largely been removed, we require an intrinsic duplex pitch of 10.55 bp/turn in order for the oxDNA N-type helix bundle to have zero global twist, and we set the pitch to this value (at $[Na^+] = 0.5$M) for oxDNA2 by modifying V_{backbone}. The requirement for a pitch of 10.55 bp/turn, rather than the 10.5 bp/turn that one might expect, is due to a few subtle effects that are currently being investigated further.

The global twist measured for the helix bundles with the new oxDNA2 model (Sec. I A of the supplementary material[69] describes how the global twist was measured from simulations) is compared to the experimental results in Table I, while typical simulation snapshots are shown in Fig. 4. The slight modifications to V_{backbone} used to set the model pitch are given in Sec. I B of the supplementary material.[69] We note that the modifications to V_{backbone} changed the duplex melting temperatures in the model by 1-2 K, which were then refitted using histogram reweighting (as described in Sec. II C 2 of Ref. 64) to give an agreement with experimental melting temperatures, which was as good as for the original model.

VI. AA/TT SEQUENCE DEPENDENCE

As a further improvement, the oxDNA2 model incorporates a more realistic sequence-dependent stacking interaction, which is achieved by differentiating between the AA and TT stacking interaction strengths. In the previous parameterisation of oxDNA,[60] the sequence-dependent base-pairing and stacking interaction strengths were obtained by fitting the oxDNA duplex melting temperatures to the SantaLucia model, a nearest-neighbor model that is able to predict experimental duplex melting temperatures very well.[49] The SantaLucia model is designed at the level of base-pair steps, where each base-pair step consists of four bases, with a free-energy difference between its single-stranded state and its duplex state. In total, there are 10 unique base-pair steps in the SantaLucia model: AA/TT, AT/AT, TA/TA, GC/GC, CG/CG, GG/CC, GA/TC, AG/CT, TG/CA, and GT/AC (for example, AG/CT refers to a base-pair step with complementary bases AG on one strand and CT on the other, both specified in the 3′ to 5′ direction). Therefore, fitting oxDNA to the SantaLucia model only allows one to find the sum of the strengths of the two stacking interactions between the nucleotides that are within a base-pair step. For example, it is only possible to find the average of the AA and TT stacking strengths from the SantaLucia model.

However, experimental evidence from sequence-dependent measurements of the mechanical properties of single-stranded DNA,[47,73] as well as hairpin stabilities and closing rates,[74,75] has revealed that sequences of A bases are much stiffer than equal length sequences of T bases. It has been argued that this is evidence that consecutive AA bases stack much more strongly than TT bases do. Here, we use original experimental data on the stabilities of hairpins with either a poly-A or a poly-T loop, to differentiate between the AA and TT stacking strengths. All hairpins have 6-bp stems and loops of either 21 or 31 bases. Details of the experimental systems and the sequences are given in Sec. III of the supplementary material.[69]

To parameterise our model to reproduce the experimental data, our procedure is to vary the AA stacking interaction strength, ϵ_{AA}, while fixing the sum of the AA and TT stacking interaction strengths to the value that is obtained by fitting oxDNA's duplex melting temperatures to the SantaLucia model, i.e.,

$$\epsilon_{\text{AA}} + \epsilon_{\text{TT}} = 2\epsilon_{\text{avg}}, \tag{8}$$

where ϵ_{TT} is the TT stacking strength, and ϵ_{avg} is the strength for both AA and TT obtained from fitting oxDNA to the SantaLucia model. We vary the AA stacking interaction strength to match the experimental differences of the thermal stabilities of the poly-A-loop and poly-T-loop hairpins, specifically $\delta\Delta F = \Delta F_{(\text{A-loop})} - \Delta F_{(\text{T-loop})}$, where $\Delta F_{(\text{X-loop})} = F_{\text{b},(\text{X-loop})} - F_{\text{ub},(\text{X-loop})}$ is the difference in free energy between the bound (hairpin) state and the unbound (open) state for a hairpin with a poly-X loop.

We obtain the bound and unbound free energies F_α, where $\alpha \in \{\text{b},\text{ub}\}$, by using thermodynamic integration. We start by calculating $F_\alpha^{(0)}$, the free energy of a reference state for which the effective charge on the backbone site, q_{eff}, is set to 0, and $\epsilon_{\text{AA}} = \epsilon_{\text{TT}} = \epsilon_{\text{avg}}$. $F_\alpha(I, x, q_{\text{eff}})$, the free energy of the state with salt concentration I, stacking strengths ϵ_{AA} and ϵ_{TT}, and effective backbone charge q_{eff}, is then obtained by solving the following integrals:

$$F_\alpha(I, x, q_{\text{eff}}) = F_\alpha^{(0)} + \int_0^{q_{\text{eff}}} \mathrm{d}q'_{\text{eff}} \frac{\partial F_\alpha(I, 0, q'_{\text{eff}})}{\partial q'_{\text{eff}}} + \int_0^x \mathrm{d}x' \frac{\partial F_\alpha(I, x', q_{\text{eff}})}{\partial x'}, \tag{9}$$

where x measures the deviation of the AA and TT strengths from ϵ_{avg}, such that Eq. (8) is satisfied, i.e.,

$$\epsilon_{\text{AA}} = (1 + x)\epsilon_{\text{avg}}, \quad \epsilon_{\text{TT}} = (1 - x)\epsilon_{\text{avg}}. \tag{10}$$

Taking the derivatives of the free energy in Eq. (9), we obtain

$$F_\alpha(I, x, q_{\text{eff}}) = F_\alpha^{(0)} + \int_0^{q_{\text{eff}}} \mathrm{d}q'_{\text{eff}} \frac{2\langle V_{\text{DH}}(I, 0, q'_{\text{eff}})\rangle_\alpha}{q'_{\text{eff}}} + \int_0^x \mathrm{d}x' \left[\frac{1}{1 + x'} \langle V_{\text{stack}}(I, x', q_{\text{eff}})\rangle_\alpha^{\text{AA}} - \frac{1}{1 - x'} \langle V_{\text{stack}}(I, x', q_{\text{eff}})\rangle_\alpha^{\text{TT}} \right], \tag{11}$$

where $\langle V_{\text{stack}}(I, x, q_{\text{eff}})\rangle_\alpha^{\text{AA(TT)}}$ is the average stacking energy of all AA(TT) nucleotides in a strand in state α with salt concentration I, x defined as in Eq. (10), and backbone charge q_{eff}. The terms appearing in the integrands of Eq. (11) can be obtained by running short simulations of bound and unbound

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22

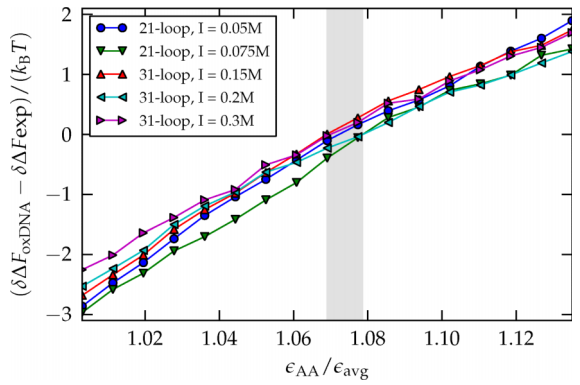

FIG. 5. Deviation of oxDNA's value for $\delta\Delta F$ from experiment at 295.6 K, as a function of $\epsilon_{\mathrm{AA}}/\epsilon_{\mathrm{avg}}$. The grey region contains the zero-deviation points for every curve. The experimental results are shown in Fig. S5, and details of the experimental setup can be found in Sec. III of the supplementary material.[69]

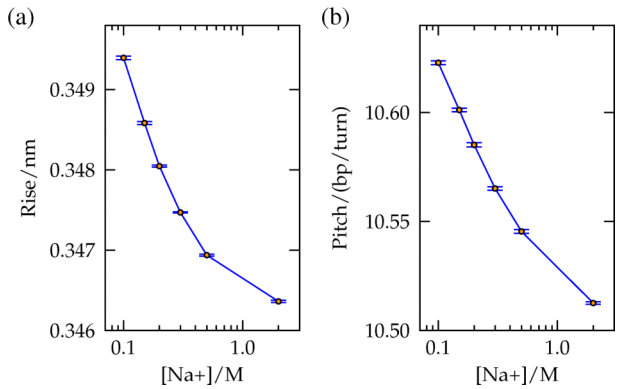

FIG. 6. (a) Rise and (b) pitch of a 60-bp duplex as a function of salt concentration in oxDNA2. For each salt concentration, the duplex was simulated for at least 3×10^9 MD steps. The error bars, which are narrower than the plot markers, show the standard error on the mean given by averaging over 10 independent estimates for each data point.

states separately and therefore the free energies can be calculated in an efficient manner. The results for the difference in $\delta\Delta F$ between oxDNA and experiment as a function of ϵ_{AA} for the hairpins with 21- and 31-base loops are shown in Fig. 5; the value for the AA stacking strength that minimises this difference is found to be $\epsilon_{\mathrm{AA}} \approx 1.075\epsilon_{\mathrm{avg}}$ (corresponding to $\epsilon_{\mathrm{TT}} \approx 0.925\epsilon_{\mathrm{avg}}$), which is not too dissimilar to the preliminary value suggested in Ref. 60. We note that this value gives satisfactory predictions for $\delta\Delta F$ for a wide range of salt concentrations down to 0.05M.

VII. THE oxDNA2 MODEL

In summary, the oxDNA2 potential can be written as

$$V_{\mathrm{oxDNA2}} = \sum_{\substack{\text{nearest}\\ \text{neighbours}}} (V^*_{\mathrm{backbone}} + V^*_{\mathrm{stack}} + V'_{\mathrm{exc}}) + \sum_{\substack{\text{other}\\ \text{pairs}}} (V^*_{\mathrm{HB}} + V_{\mathrm{cross\ stack}} + V_{\mathrm{exc}} + V^*_{\mathrm{coax\ stack}} + V^*_{\mathrm{DH}}), \tag{12}$$

where a V^*_{x} indicates that the term is either modified for oxDNA2 or, in the case of V^*_{DH}, new in oxDNA2. The modified parameters for oxDNA2 are compared with those for oxDNA in Table S1, and a full account of the changes is given in Sec. I of the supplementary material.[69] All other parameters remain the same as in the original model.

We emphasise that, after all the relevant changes to the potential were made, the hydrogen bonding and stacking parameters were modified to ensure that the close agreement to experimental duplex melting temperatures achieved for the original model was retained with oxDNA2 (this was done immediately after the changes to V_{backbone}, as described at the end of Sec. V).

VIII. PHYSICAL PROPERTIES OF oxDNA2

The structural, mechanical, and thermodynamic properties of DNA for the new version of oxDNA presented in this paper, which we call oxDNA2, are slightly different from the properties for the original oxDNA model. We briefly highlight the most important of these changes here; in addition, these properties are given as a function of salt concentration as this is now possible with the new model. We note that all of the results described in this section were computed using the final version of the oxDNA2 model as summarized in Sec. VII. The details of the simulations used to compute these results are given briefly in the figure captions and in detail in Sec. II C of the supplementary material and Table S3.[69]

The structural properties of the new model, specifically the pitch and rise of double-stranded DNA, are presented in Fig. 6. As might be expected, the rise increases with decreasing salt concentration, due to the greater repulsion between backbone sites. The pitch also increases with decreasing salt, consistent with the measured increase in rise and slight decrease in neighbouring backbone-backbone distance measured as salt concentration is decreased (the duplex radius remains approximately constant). Although there is not much experimental evidence to compare this with, there is some indication that the pitch is roughly constant for low salt concentrations (0.162M and below).[50] As mentioned earlier, the pitch is chosen (by modifying the bonded neighbour backbone-backbone interaction) so that the global twist of origami structures agrees with experimental measurements. Specifically, we set the backbone-backbone interaction so that the helix bundle designed to have no global twist has no global twist in the model at $[\mathrm{Na}^+] = 0.5$M. This results in a pitch of roughly 10.55 bp/turn at this salt concentration, compared to 10.36 bp/turn in the original model, and experimental values of around 10.45 suggested by cyclisation experiments,[50,53] albeit in the presence of some divalent salt.

The thermodynamics of duplex formation are shown in Fig. 7. The transition width for the yield curve of a 10-bp duplex at 0.5M $[\mathrm{Na}^+]$ in oxDNA2 is largely unchanged from the original oxDNA, and the transition widths depend weakly if at all on salt in oxDNA2. The free-energy profiles for duplex formation in oxDNA2 show that the free-energy cost of forming the first base pair decreases with increasing salt, presumably due to the reduced energetic cost of bringing the two single strands close together as the electrostatic repulsion between backbone sites becomes more short-ranged. The slope of the bound region of the free-energy profile also becomes

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22

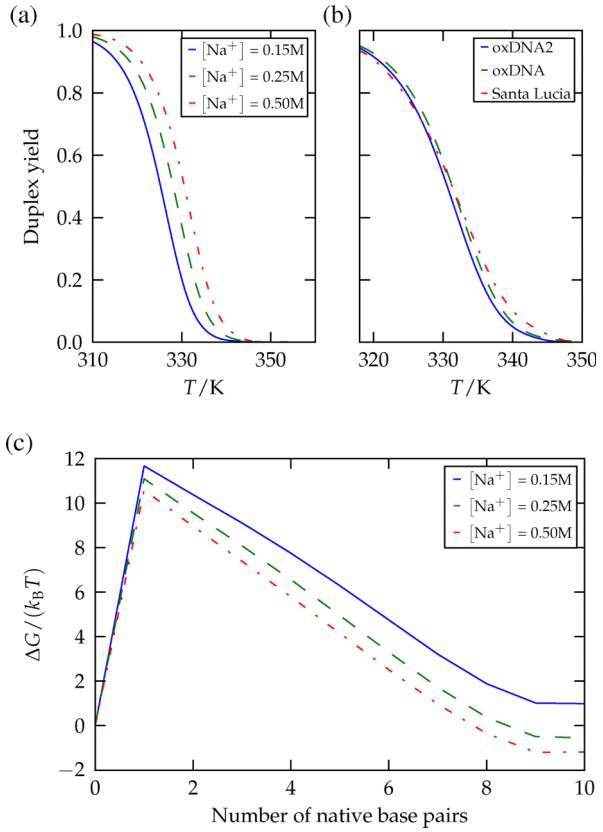

FIG. 7. The thermodynamics of formation of a 10-bp duplex with oxDNA2. (a) Duplex yield for oxDNA2 at different salt concentrations. (b) Comparison of duplex yield profiles at $[Na^+] = 0.5$M as predicted by oxDNA2, the original oxDNA model and the SantaLucia model.[49] (c) The free-energy profile for duplex formation at different salt concentrations, at the melting temperature of a 10-bp duplex at $[Na^+] = 0.25$M. The duplex was simulated for at least 3×10^{10} virtual move Monte Carlo[76] (VMMC) steps at each salt concentration using umbrella sampling at a monomer concentration of 3.3×10^{-4}M.

steeper with increasing salt, indicating an increased free-energy gain on forming subsequent base pairs.

In Fig. 8, the duplex melting temperatures as a function of salt and for different duplex lengths in oxDNA2 are compared to the results from the SantaLucia model, to which oxDNA2 was parameterised. As expected, we find good agreement.

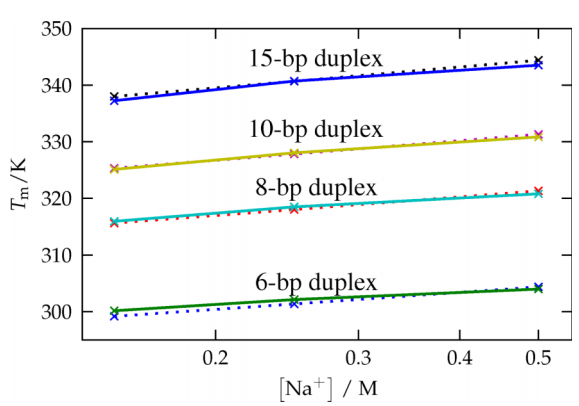

FIG. 8. The melting temperature as a function of salt concentration for duplexes of different lengths simulated with oxDNA2 (crosses with solid lines) and as predicted by the SantaLucia model (crosses with dotted lines) for a DNA strand concentration of 3.3×10^{-4}M. Each duplex was simulated for roughly 4×10^9 VMMC steps at a salt concentration of 0.25M, and the melting temperatures at 0.15M and 0.5M were computed by single histogram reweighting.[77]

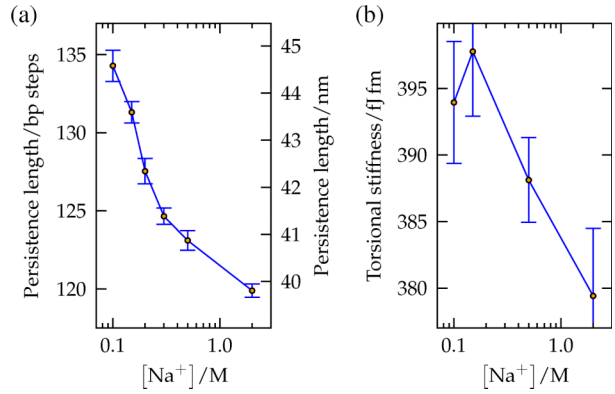

FIG. 9. (a) The persistence length and (b) the torsional stiffness of duplex DNA in oxDNA2 as a function of salt concentration. For the persistence length, a 60-bp duplex was simulated for at least 3×10^9 MD steps for each salt concentration. The method for computing the torsional stiffness is given in Sec. II B of the supplementary material.[69] Error bars for the simulation results show the standard error on the mean given by averaging over either 10 independent estimates (for the persistence length) or 5 independent estimates (for the torsional stiffness) for each data point.

The mechanical properties of the model's double-stranded DNA as a function of salt, specifically the persistence length and torsional stiffness, are shown in Fig. 9. The persistence length was calculated by computing the correlation of the helix axis as described in Ref. 30. The effective torsional stiffness, C_{eff}, was computed from the linear regime of the torque response curve of a 60-bp duplex under tension (see Sec. II B of the supplementary material[69] for details). The simulations were carried out under a linear force of 30 pN, a high force regime where we expect that C_{eff} approximates the true torsional stiffness C_0.[43,78] The persistence length at $[Na^+] = 0.5$M, about 123 bp, is very slightly lower than for the original model, 125 bp, and the persistence length at different salt concentrations is consistent with the rather broad range of values suggested by experiment.[79,80] The decrease in persistence length with increasing salt is expected due to the decrease in repulsion between the duplex's backbone sites, which makes the duplex less stiff, in agreement with experiment.[79] The slight decrease in torsional stiffness with increasing salt can be rationalised in the same way. Regardless of the salt concentration, the torsional stiffness measured for oxDNA2 ($C_{\text{eff}} \approx 380 - 400$ fJfm) is lower than for the original model[43] ($C_0 \approx 473$ fJfm). Single-molecule twisting experiments give a value of $C_{\text{eff}} \approx 410$ fJfm for a pulling force of 3.5 pN[81] and a pulling force of 45 pN,[82] at around $[Na^+] = 0.1$M. There are limited experimental data on the salt dependence of torsional stiffness in DNA; however, the torsional stiffness has been reported to be roughly constant in the range $[Na^+] = 0$-0.162M.[50]

The effects of various motifs on duplex melting temperatures in oxDNA2 are compared to results from the original oxDNA and the SantaLucia model in Table II. These effects are either barely changed in oxDNA2 or are now closer to the SantaLucia values than the original oxDNA values were.

The melting temperatures of hairpins at 0.5M $[Na^+]$ are shown in Fig. 10. The hairpin melting temperatures are lower than for the original oxDNA by about 2 K, which in turn had hairpin melting temperatures about 3 K lower than those predicted by SantaLucia. The gap between the oxDNA2 hairpin

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22

TABLE II. The effect of introducing various motifs on the duplex melting temperature at a monomer concentration of 3.3×10^{-4}M. ΔT_m is the difference between the melting temperatures of the structure with the motif and a duplex consisting of the same number of complementary base pairs as the motif structure. For bulges and internal mismatches, the motif was placed at the centre of the duplex. The simulations of the bulges, terminal mismatches and internal mismatches were run for at least 6×10^{10}, 8.5×10^{10}, and 5.6×10^{10} VMMC steps, respectively. All oxDNA2 results were obtained from simulations at a concentration of 4.2×10^{-5}M which were extrapolated to a concentration of 3.3×10^{-4}M.

	ΔT_m/K		
Motif	oxDNA2	oxDNA	SantaLucia
Bulge (1-base bulge in 8-bp duplex)	−18.7	−17.98	−23.19
Bulge (2-base bulge in 8-bp duplex)	−27.4	−23.92	−26.73
Terminal mismatch (1-bp mismatch in 5-bp duplex)	+6.5	+6.71	+8.6
Internal mismatch (2-bp mismatch in 8-bp duplex)	−15.9	−15.77	−13.99

melting temperatures and those given by the SantaLucia model widens somewhat as the salt concentration is lowered (Fig. 11), indicating a stronger salt dependence in oxDNA2, with a typical underestimate of around 7 K at a salt concentration of 100 mM for a relatively short loop length of 6 bases. For longer loops, the difference in predicted melting temperatures between oxDNA2 and SantaLucia widens further. This difference is unsurprising, as in oxDNA2 (and physical DNA), ssDNA becomes stiffer at lower salt concentrations, making the formation of a hairpin less favorable,[44] whereas in the SantaLucia model, the loops' contribution to hairpin stability is salt-independent.[49]

It seems plausible that oxDNA2's performance is better than implied by the salt-independent loop contribution in the SantaLucia model. In Fig. 12, we consider hairpins with a short 5-bp stem and a long 31-base loop, making the hairpin thermodynamics particularly sensitive to the change in loop stiffness with salt. The hairpin stability (i.e., the free-energy difference between the bound and unbound states) as predicted by oxDNA2 and the SantaLucia model is compared to experimental data determined using Förster resonance energy transfer (FRET) measurements (experimental details in Sec. III of the supplementary material[69]). As expected, we find the hairpin stabilities for oxDNA2 have a much stronger salt dependence (steeper gradient) than those of SantaLucia (shallow gradient). At higher salt, the oxDNA2 results show a similar slope to the experimental curve, but at lower salt

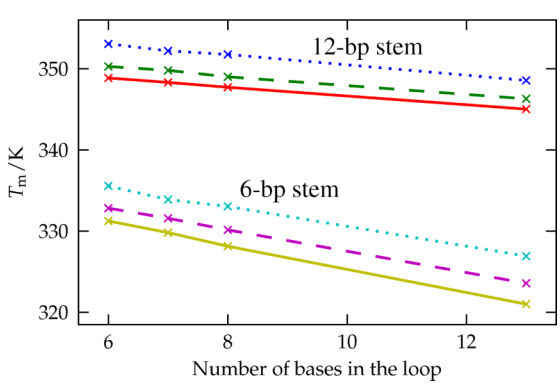

FIG. 10. The melting temperature as a function of loop length for hairpins simulated with oxDNA2 (crosses with solid lines), the original oxDNA (crosses with dashed lines), and according to the SantaLucia model (crosses with dotted lines), at 0.5M [Na^+]. For the oxDNA2 results, the hairpins were simulated at each salt for at least 1.8×10^{10} VMMC steps.

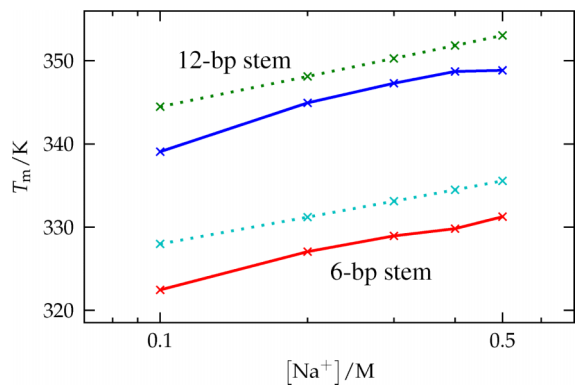

FIG. 11. The melting temperature as a function of salt concentration of hairpins with 6-base loops simulated with oxDNA2 (crosses with solid lines) and according to the SantaLucia model (crosses with dotted lines). For the oxDNA2 results, the hairpins were simulated at each salt for at least 1.5×10^{10} VMMC steps.

(0.2M and below), oxDNA2 shows a stronger destabilization of the hairpins with decreasing salt (a steeper gradient) than experiment. Thus, although oxDNA2 does a reasonable job of capturing these physical effects, the comparison suggests that the single strands may be experiencing too much repulsion at

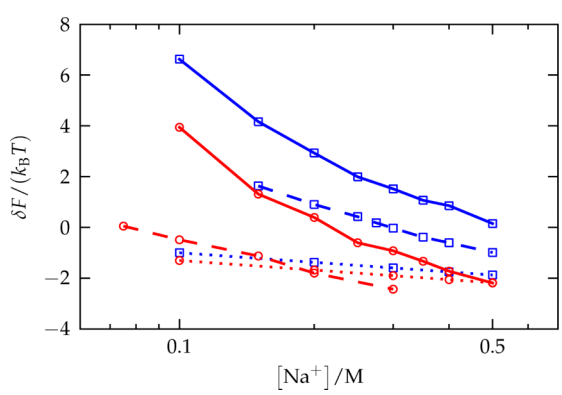

FIG. 12. The free energy difference between the bound and unbound state for two hairpins as measured in experiment using FRET (dashed lines), and as predicted by oxDNA2 (solid lines) and the SantaLucia model (dotted lines). Results for two different hairpins are presented: the A_{31}-VW (red, circular markers) hairpin and the T_{31}-VW hairpin (blue, square markers). The sequences for these hairpins are given in Table S4.[69] The oxDNA2 simulations were run for 10^{10} VMMC steps for each hairpin, with $q_{eff} = 0$ and $x = 0$ (with x defined in Eq. (10)), and the results (computed for the standard oxDNA2 values for q_{eff} and x) were calculated using thermodynamic integration as described in Sec. VI and Sec. II C of the supplementary material.[69]

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22

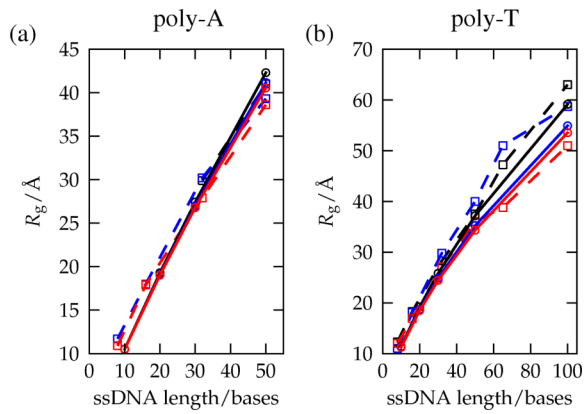

FIG. 13. The radius of gyration (R_g) for ssDNA as a function of ssDNA length and for different salt concentrations, for (a) a poly-A ssDNA and (b) a poly-T ssDNA, from oxDNA2 (solid lines with circles) and from experiments due to Sim *et al.*[83] (dashed lines with squares). The salt concentrations shown are 0.225M (black lines), 0.525M (blue lines), and 1.025M (red lines). For the oxDNA2 results, the strands were simulated for between 2×10^9 and 5×10^{10} VMMC steps at each strand length and salt concentration.

low salt. We also note that the SantaLucia model is unable to predict the large difference in stability of the two hairpins due to their different loop sequences, because the SantaLucia model is insensitive to the loop sequence except for the two bases adjacent to the stem.

We examine the flexibility of single strands as a function of salt in oxDNA2 by measuring the radius of gyration, R_g, of a single strand of DNA at different salt concentrations, sequences, and strand lengths and comparing to experimental results (Fig. 13). The oxDNA2 model reproduces the overall trend of increasing R_g with salt and is in agreement with the somewhat noisy experimental data. We note that oxDNA2 is able to capture two important effects: the greater stiffness (and hence larger R_g) of the more strongly stacking poly-A strands and the greater salt dependence of the poly-T strands' R_g compared to that of the poly-A strands, caused by the weaker stacking of the poly-T strands, which means that electrostatic repulsion makes a greater relative contribution to its stiffness.

IX. CONCLUSION

oxDNA has been applied to a wider variety of systems than any other coarse-grained model of DNA. The modifications and extensions to oxDNA presented here open up a variety of new potential applications for the model. With the introduction of an explicit salt-dependent term in the potential, the model can be used to simulate systems under physiological conditions and to investigate the salt-dependent behaviour of DNA. Also, the parametrization presented here allows a quantitative comparison to experiments run in a wide range of salt concentrations rather than just in the high-salt limit. The introduction of major-minor grooving adds detail to the model and, combined with other small modifications, allows the use of oxDNA2 to accurately characterise the structural properties of large DNA nanostructures, such as DNA origami. In particular, the helical pitch and the twist angles at nicks and junctions were fine-tuned to obtain a correct global twist for a test-case 3D origami structure. In the absence of definitive experimental values for these individual parameters, we chose a combination that we deemed physically reasonable and that produces equilibrium structures that compare well with experimental ones. Finally, the sequence dependence in the model has been extended by introducing different interaction strengths for the AA and TT stacking. This change will be particularly useful for studying the effects of stacking in single strands or single-stranded sections (e.g., hairpin loops) as poly-A and poly-T sections are often used as paradigmatic examples of strongly and weakly stacking sequences, respectively, and for modelling DNA nanostructures where poly-T loops are often used as flexible linkers.

There are some limitations associated with the new features of the model. The Debye-Hückel treatment of the electrostatics is perhaps an oversimplification, and we should not expect it to capture all of the complexities associated with the electrostatics for DNA, but it is not straightforward to think of a different approach that would be consistent with our level of coarse graining. In particular, the fact that hairpin stability is reduced faster than in experiment suggests that single strands may experience too much repulsion in oxDNA2 at low salt, and care should be taken in making predictions based on our models at low salt concentration if the system under investigation depends crucially on the thermodynamics of long single-stranded sections.

We are currently exploiting the improvements in the model's structural prediction to carry out investigations on DNA origami structures as well as on structures composed of multi-arm tiles.[84] At the same time, we are studying the salt dependent thermodynamics of a diverse set of systems, which would not have been possible without the introduction of salt dependence into the potential. The introduction of different strengths for the AA and TT stacking in the model allows us to better capture these effects and in addition gives us more accurate sequence-dependent hairpin thermodynamics and kinetics, which are currently being exploited to further study hairpins. We note that a simulation code implementing both the original oxDNA model and the new oxDNA2 model, including an implementation using graphical processing units (GPUs),[85] is available for download from dna.physics.ox.ac.uk.

ACKNOWLEDGMENTS

The authors are grateful to the Engineering and Physical Sciences Research Council for financial support. M.M. was supported by the Swiss National Science Foundation (Grant No. PBEZP2-145981). The authors thank Oxford's Advanced Research Computing, E-infrastructure South, and the PolyHub virtual organization for computer time.

[1]J. Zheng, J. J. Birktoft, Y. Chen, T. Wang, R. Sha, P. E. Constantinou, S. L. Ginell, C. Mao, and N. C. Seeman, Nature **461**, 74 (2009).
[2]H. Dietz, S. M. Douglas, and W. M. Shih, Science **325**, 725 (2009).
[3]D. Han, S. Pal, Y. Yang, S. Jiang, J. Nangreave, Y. Liu, and H. Yan, Science **339**, 1412 (2013).
[4]Y. Ke, L. L. Ong, W. M. Shih, and P. Yin, Science **338**, 1177 (2012).
[5]H. K. K. Subramanian, B. Chakraborty, R. Sha, and N. C. Seeman, Nano Lett. **11**, 910 (2011).
[6]A. S. Walsh, H. Yin, C. M. Erben, M. J. A. Wood, and A. J. Turberfield, ACS Nano **5**, 5427 (2011).
[7]S. M. Douglas, I. Bachelet, and G. M. Church, Science **335**, 831 (2012).
[8]A. Kuzyk, R. Schreiber, Z. Fan, G. Pardatscher, E.-M. Roller, A. Hogele, F. C. Simmel, A. O. Govorov, and T. Liedl, Nature **483**, 311 (2012).

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22

[9]R. M. Zadegan, M. D. E. Jepsen, K. E. Thomsen, A. H. Okholm, D. H. Schaffert, E. S. Andersen, V. Birkedal, and J. Kjems, ACS Nano **6**, 10050 (2012).
[10]A. V. Pinheiro, D. Han, W. M. Shih, and H. Yan, Nat. Nanotechnol. **6**, 763 (2011).
[11]D. Svozil, P. Hobza, and J. Šponer, J. Phys. Chem. B **114**, 1191 (2010).
[12]A. Mladek, M. Krepl, D. Svozil, P. Cech, M. Otyepka, P. Banas, M. Zgarbova, P. Jurecka, and J. Šponer, Phys. Chem. Chem. Phys. **15**, 7295 (2013).
[13]J. Šponer, J. E. Šponer, A. Mldek, P. Ban, P. Jureka, and M. Otyepka, Methods **64**, 3 (2013).
[14]C. A. Laughton and S. A. Harris, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **1**, 590 (2011).
[15]A. Pérez, F. J. Luque, and M. Orozco, Acc. Chem. Res. **45**, 196 (2012).
[16]C. Maffeo, B. Luan, and A. Aksimentiev, Nucleic Acids Res. **40**, 3812 (2012).
[17]J. Yoo and A. Aksimentiev, Proc. Natl. Acad. Sci. U. S. A. **110**, 20099 (2013).
[18]C. Maffeo, J. Yoo, J. Comer, D. B. Wells, B. Luan, and A. Aksimentiev, J. Phys.: Condens. Matter **26**, 413101 (2014).
[19]C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, Science **265**, 1599 (1994).
[20]D. M. Hinckley, G. S. Freeman, J. K. Whitmer, and J. J. de Pablo, J. Chem. Phys. **139**, 144903 (2013).
[21]A. Savelyev and G. A. Papoian, Proc. Natl. Acad. Sci. U. S. A. **107**, 20340 (2010).
[22]T. Cragnolini, P. Derreumaux, and S. Pasquali, J. Phys. Chem. B **117**, 8047 (2013).
[23]Y. He, M. Maciejczyk, S. Ołdziej, H. A. Scheraga, and A. Liwo, Phys. Rev. Lett. **110**, 098101 (2013).
[24]M. C. Linak, R. Tourdot, and K. D. Dorfman, J. Chem. Phys. **135**, 205102 (2011).
[25]A. Morriss-Andrews, J. Rottler, and S. S. Plotkin, J. Chem. Phys. **132**, 035105 (2010).
[26]J. C. Araque, A. Z. Panagiotopoulos, and M. A. Robert, J. Chem. Phys. **134**, 165103 (2011).
[27]N. Korolev, D. Luo, A. P. Lyubartsev, and L. Nordenskiöld, Polymers **6**, 1655 (2014).
[28]C. Maffeo, T. T. M. Ngo, T. Ha, and A. Aksimentiev, J. Chem. Theory Comput. **10**, 2891 (2014).
[29]O. Gonzalez, D. Petkevičiūtė, and J. Maddocks, J. Chem. Phys. **138**, 055102 (2013).
[30]T. E. Ouldridge, A. A. Louis, and J. Doye, J. Chem. Phys. **134**, 085101 (2011).
[31]T. E. Ouldridge, Ph.D. thesis, University of Oxford, 2011.
[32]T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, Phys. Rev. Lett. **104**, 178101 (2010).
[33]P. Šulc, T. E. Ouldridge, F. Romano, J. P. K. Doye, and A. A. Louis, Nat. Comput. **13**, 535 (2014).
[34]T. E. Ouldridge, P. Šulc, F. Romano, J. P. K. Doye, and A. A. Louis, Nucleic Acids Res. **41**, 8886 (2013).
[35]T. E. Ouldridge, R. L. Hoare, A. A. Louis, J. P. K. Doye, J. Bath, and A. J. Turberfield, ACS Nano **7**, 2479 (2013).
[36]L. Rovigatti, F. Smallenburg, F. Romano, and F. Sciortino, ACS Nano **8**, 3567 (2014).
[37]J. S. Schreck, T. E. Ouldridge, F. Romano, A. A. Louis, and J. P. K. Doye, J. Chem. Phys. **142**, 165101 (2015).
[38]J. S. Schreck, T. E. Ouldridge, F. Romano, P. Šulc, L. Shaw, A. A. Louis, and J. P. K. Doye, "DNA hairpins primarily promote duplex melting rather than inhibiting hybridization," Nucleic Acids Res. (in press); e-print arXiv:1408.4401.
[39]N. Srinivas, T. E. Ouldridge, P. Šulc, J. M. Schaeffer, B. Yurke, A. A. Louis, J. P. K. Doye, and E. Winfree, Nucleic Acids Res. **41**, 10641 (2013).
[40]M. Mosayebi, A. A. Louis, J. P. K. Doye, and T. E. Ouldridge, "Force-induced rupture of a DNA duplex," preprint arXiv:1502.03623 (2015).
[41]C. Matek, T. E. Ouldridge, A. Levy, J. P. K. Doye, and A. A. Louis, J. Phys. Chem. B **116**, 11616 (2012).
[42]Q. Wang and B. M. Pettitt, Biophys. J. **106**, 1182 (2014).
[43]C. Matek, T. E. Ouldridge, J. P. K. Doye, and A. A. Louis, Sci. Rep. **5**, 7655 (2015).
[44]M. Mosayebi, F. Romano, T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Phys. Chem. B **118**, 14326 (2014).
[45]F. Romano, D. Chakraborty, J. P. K. Doye, T. E. Ouldridge, and A. A. Louis, J. Chem. Phys. **138**, 085101 (2013).
[46]R. R. F. Machinek, T. E. Ouldridge, N. E. C. Haley, J. Bath, and A. J. Turberfield, Nat. Commun. **5**, 5324 (2014).
[47]A. Bosco, J. Camunas-Soler, and F. Ritort, Nucleic Acids Res. **42**, 2064 (2013).
[48]M. E. Polinkovsky, Y. Gambin, P. R. Banerjee, M. J. Erickstad, A. Groisman, and A. A. Deniz, Nat. Commun. **5**, 5737 (2014).
[49]J. SantaLucia, Jr. and D. Hicks, Annu. Rev. Biophys. Biomol. Struct. **33**, 415 (2004).
[50]W. H. Taylor and P. J. Hagerman, J. Mol. Biol. **212**, 363 (1990).
[51]F. Mocci and A. Laaksonen, Soft Matter **8**, 9268 (2012).
[52]D. A. Potoyan, A. Savelyev, and G. A. Papoian, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **3**, 69 (2013).
[53]D. Shore and R. L. Baldwin, J. Mol. Biol. **170**, 983 (1983).
[54]A. Dawid, F. Guillemot, C. Brème, V. Croquette, and F. Heslot, Phys. Rev. Lett. **96**, 188102 (2006).
[55]S. Ido, K. Kimura, N. Oyabu, K. Kobayashi, M. Tsukada, K. Matsushige, and H. Yamada, ACS Nano **7**, 1817 (2013).
[56]Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, and A. Vologodskii, Proc. Natl. Acad. Sci. U. S. A. **102**, 5397 (2005).
[57]J. C. Wang, Proc. Natl. Acad. Sci. U. S. A. **76**, 200 (1979).
[58]T. Kato, R. P. Goodman, C. M. Erben, A. J. Turberfield, and K. Namba, Nano Lett. **9**, 2747 (2009).
[59]X.-C. Bai, T. G. Martin, S. H. W. Scheres, and H. Dietz, Proc. Natl. Acad. Sci. U. S. A. **109**, 20012 (2012).
[60]P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. K. Doye, and A. A. Louis, J. Chem. Phys. **137**, 135101 (2012).
[61]F. F. Andersen, B. Knudsen, C. L. P. Oliveira, R. F. Frohlich, D. Kruger, J. Bungert, M. Agbandje-McKenna, R. McKenna, S. Juul, C. Veigaard, J. Koch, J. L. Rubinstein, B. Guldbrandtsen, M. S. Hede, G. Karlsson, A. H. Andersen, J. S. Pedersen, and B. R. Knudsen, Nucleic Acids Res. **36**, 1113 (2008).
[62]C. Tian, C. Zhang, X. Li, C. Hao, S. Ye, and C. Mao, Langmuir **30**, 5859 (2013).
[63]S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, and W. M. Shih, Nature **459**, 414 (2009).
[64]P. Šulc, F. Romano, T. E. Ouldridge, J. P. K. Doye, and A. A. Louis, J. Chem. Phys. **140**, 235102 (2014).
[65]C. Zhang, W. Wu, X. Li, C. Tian, H. Qian, G. Wang, W. Jiang, and C. Mao, Angew. Chem., Int. Ed. **51**, 7999 (2012).
[66]D. M. Hinckley, J. P. Lequieu, and J. J. de Pablo, J. Chem. Phys. **141**, 035102 (2014).
[67]C. Maffeo, R. Schöpflin, H. Brutzer, R. Stehr, A. Aksimentiev, G. Wedemann, and R. Seidel, Phys. Rev. Lett. **105**, 158101 (2010).
[68]J. Mazur and R. L. Jernigan, Biopolymers **31**, 1615 (1991).
[69]See supplementary material at http://dx.doi.org/10.1063/1.4921957 for further details of methods and results.
[70]D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, 1996).
[71]C. Vega, E. Sanz, J. Abascal, and E. Noya, J. Phys.: Condens. Matter **20**, 153101 (2008).
[72]T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Phys.: Condens. Matter **22**, 104102 (2010).
[73]T. Uzawa, T. Isoshima, Y. Ito, K. Ishimori, D. E. Makarov, and K. W. Plaxco, Biophys. J. **104**, 2485 (2013).
[74]N. L. Goddard, G. Bonnet, O. Krichevsky, and A. Libchaber, Phys. Rev. Lett. **85**, 2400 (2000).
[75]G. Bonnet, O. Krichevsky, and A. Libchaber, Proc. Natl. Acad. Sci. U. S. A. **95**, 8602 (1998).
[76]S. Whitelam, E. H. Feng, M. F. Hagan, and P. L. Geissler, Soft Matter **5**, 1251 (2009).
[77]S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **13**, 1011 (1992).
[78]J. D. Moroz and P. Nelson, Proc. Natl. Acad. Sci. U. S. A. **94**, 14418 (1997).
[79]A. Savelyev, Phys. Chem. Chem. Phys. **14**, 2250 (2012).
[80]E. Herrero-Galn, M. E. Fuentes-Perez, C. Carrasco, J. M. Valpuesta, J. L. Carrascosa, F. Moreno-Herrero, and J. R. Arias-Gonzalez, J. Am. Chem. Soc. **135**, 122 (2013).
[81]X. J. A. Janssen, J. Lipfert, T. Jager, R. Daudey, J. Beekman, and N. H. Dekker, Nano Lett. **12**, 3634 (2012).
[82]Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, Nature **424**, 338 (2003).
[83]A. Y. L. Sim, J. Lipfert, D. Herschlag, and S. Doniach, Phys. Rev. E **86**, 021901 (2012).
[84]C. Zhang, Y. He, M. Su, S. H. Ko, T. Ye, Y. Leng, X. Sun, A. E. Ribbe, W. Jiang, and C. Mao, Faraday Discuss. **143**, 221 (2009).
[85]L. Rovigatti, P. Šulc, I. Z. Reguly, and F. Romano, J. Comput. Chem. **36**, 1 (2015).

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 155.198.12.107 On: Fri, 10 Jul 2015 13:09:22