A Primer on the oxDNA Model of DNA: When to Use it, How to Simulate it and How to Interpret the Results

A. Sengar[1], T. E. Ouldridge[1]*, O. Henrich[2], L. Rovigatti[3,4] and P. Šulc[5,6]

[1]Centre for Synthetic Biology, Department of Bioengineering, Imperial College London, London, United Kingdom, [2]Department of Physics, SUPA, University of Strathclyde, Glasgow, United Kingdom, [3]Department of Physics, Sapienza University of Rome, Rome, Italy, [4]CNR Institute of Complex Systems, Sapienza University of Rome, Rome, Italy, [5]Center for Molecular Design and Biomimetics, The Biodesign Institute, Arizona State University, Tempe, AZ, United States, [6]School of Molecular Sciences, Arizona State University, Tempe, AZ, United States

The oxDNA model of Deoxyribonucleic acid has been applied widely to systems in biology, biophysics and nanotechnology. It is currently available via two independent open source packages. Here we present a set of clearly documented exemplar simulations that simultaneously provide both an introduction to simulating the model, and a review of the model's fundamental properties. We outline how simulation results can be interpreted in terms of—and feed into our understanding of—less detailed models that operate at larger length scales, and provide guidance on whether simulating a system with oxDNA is worthwhile.

1 INTRODUCTION

Deoxyribonucleic acid (DNA) is a macromolecule that acts as a storage medium for genetic information for all living organisms (Alberts et al., 2002). In nature, the molecule is most often found as a double helix of two strands. The structure of each strand comprises of a backbone of covalently linked sugar and phosphate groups. Each sugar is further attached to a base moiety: adenine (A), guanine (G), cytosine (C) or thymine (T). Certain intra- and intermolecular interactions between these bases drive the formation of the aforementioned double helical structure.

Crucially, the base pairing that holds these duplexes together is highly specific; to a first approximation, A will only bind to T and C will only bind to G, and vice versa. Matching—or complementary—sequences therefore bind to each other much more strongly than to non-complementary sequences. The different base identities, along with the rules of complementarity, allow information to be encoded into the single strands and copied from generation to generation (Watson and Crick, 1953).

The DNA double helix has a diameter of about 2 nm, and a helical pitch of about 3.4–3.6 nm. Double strands are relatively stiff, with large bending disfavoured on lengthscales below around 40–50 nm (Seeman, 2003). By contrast, single strands are very flexible (Murphy et al., 2004; Chen et al., 2012) forming loops and kinks with only a handful of bases or fewer.

These thermodynamic, mechanical and structural properties influence DNA's biological role, but also make it an ideal material for nanoscale engineering. The simplicity of interactions between strands, and the predictability of the structural and mechanical properties of the product, have enabled the rational design of a host of synthetic structures (Fu and Seeman, 1993; Goodman et al., 2005; Rothemund, 2006; Douglas et al., 2009; Ke et al., 2012; Zhang et al., 2015; Tikhomirov et al., 2017; Wagenbauer et al., 2017), computing architectures (Adleman, 1994; Rothemund et al., 2004;

Qian et al., 2011; Cherry and Qian, 2018; Woods et al., 2019) and dynamic systems (Yurke et al., 2000; Shin and Pierce, 2004; Muscat et al., 2011; Zhang and Seelig, 2011; Wickham et al., 2012; Srinivas et al., 2017; Tomov et al., 2017).

DNA's importance to biology, nanotechnology and simply as a canonical model biopolymer for biophysicists means that modelling its behaviour is a key challenge. Unsurprisingly, therefore, models spanning an enormous range of complexity have been proposed to analyse and rationalize the behaviour of DNA. In this pedagogical review, we will first discuss this range of models and their interplay, before focusing on a particular coarse-grained model, oxDNA.

The oxDNA model, first published in 2010 (Ouldridge et al., 2010a) (and with a slightly updated potential in 2011 (Ouldridge et al., 2011)), has now been extensively applied to problems in nanotechnology (Ouldridge et al., 2013a; Doye et al., 2013; Srinivas et al., 2013; Machinek et al., 2014; Snodin et al., 2016, 2019; Henning-Knechtel et al., 2017; Hong et al., 2018), soft matter (De Michele et al., 2012; Rovigatti et al., 2014; Procyk et al., 2020; Stoev et al., 2020), biophysics (Matek et al., 2012; Romano et al., 2013; Matek et al., 2015; Mosayebi et al., 2015; Harrison et al., 2019; Nomidis et al., 2019) and biology (Lee et al., 2015; Wang et al., 2015; Craggs et al., 2019). Numerous tools exist to generate and visualize systems with oxDNA (Henrich et al., 2018; Suma et al., 2019), alongside two independent, publicly-available code bases for actually running simulations with at least three qualitatively distinct algorithms for simulating the model (Ouldridge et al., 2011; Snodin et al., 2015). One of these code bases has recently been incorporated into a webserver (Poppleton et al., 2020).

Despite this uptake, however, there is insufficient clarity on how the basic properties of the oxDNA model make it well- or poorly-suited to studying certain systems. Moreover, many interesting phenomena require non-trivial simulation techniques if they are to be probed with oxDNA. Although those techniques have been widely applied, and software implementing them with oxDNA is available, documentation supporting their use is limited. Equally, there is very little help with the intuition required to use these techniques successfully. Finally, a major aspect to interpreting the results from oxDNA is rationalizing its predictions in terms of less detailed models. Unfortunately, however, there are many subtleties in doing so.

In this pedagogical review we implement a series of exemplar simulations that allow us to address these shortcomings. These simulations will establish a well-documented set of examples for a series of approaches that can be adapted by users, and this review will provide some of the intuition for how to use these approaches successfully. Simultaneously, we will use these examples to illustrate key aspects of the oxDNA model that determine its usefulness, and will explore how to interpret the results in terms of DNA models at different scales.

2 DNA MODELS ACROSS LENGTH SCALES

At the smallest and most fundamental scale, quantum chemistry calculations can be used to estimate the nucleotide properties from first principles (Hobza and Šponer, 1999; Pérez et al., 2004;

Šponer et al., 2004; Šponer et al., 2008). However, these calculations are computationally extremely expensive and are unable to capture the collective behaviour of whole strands in solution. Nonetheless, insight from this field has been incorporated into classical atomistic force fields AMBER (Cornell et al., 1996) and CHARMM (Brooks et al., 1983) that use empirical force fields to model interactions between atoms. These force fields are iteratively parameterised using both comparison to experimental data and information from lower-level quantum mechanical descriptions. In recent years, advances in computational resources have allowed these models to simulate large systems—such as DNA origami—for long enough timescales to analyse their equilibrium properties. Given long simulations, these atomistic models are able to sample the conformation of large structures (Nguyen et al., 2014; Rocklin et al., 2017) and the breaking and formation of base pairs (Brown et al., 2015). However, at the time of writing, a systematic study of DNA duplex formation thermodynamics, as represented by atomistic models, has not been performed. As such, it is unknown how well these atomistic models represent DNA thermodynamics—historically, the force fields have required adjustment as new systems and longer time scales are studied (Pérez et al., 2007; Yoo and Aksimentiev, 2012). This fact, alongside the heavy computational load in simulating large systems or significant structural changes, mean that atomistic approaches are currently limited to a fraction of the systems of interest in DNA-based biophysics, biology, soft matter and nanotechnology.

In an effort to access longer timescales, a number of "coarse-grained" or "mesoscale" models have been introduced (Savelyev and Papoian, 2009; Ouldridge et al., 2011; Hinckley et al., 2013; Korolev et al., 2014; Maciejczyk et al., 2014; Maffeo et al., 2014; Machado and Pantano, 2015; Uusitalo et al., 2015; Dans et al., 2016; Ivani et al., 2016; Chakraborty et al., 2018; Maffeo and Aksimentiev, 2020). These models represent DNA with a much-reduced set of degrees of freedom relative to atomistic approaches. In particular, solvent (and solvated ions) are usually treated implicitly, and groups of atoms in the DNA are replaced by a single site with effective interactions. As a result, these models can access longer length and time scales than atomistic descriptions.

The procedure for coarse-graining ranges from "bottom-up" approaches that seek to formally map the statistical behaviour of a more detailed model into a coarse-grained description (Savelyev and Papoian, 2009; Maciejczyk et al., 2014; Maffeo et al., 2014), to "top-down" approaches such as oxDNA that are more *ad hoc*, instead seeking to reproduce as many experimentally relevant properties as possible (Ouldridge et al., 2009; Hinckley et al., 2013; Machado and Pantano, 2015; Uusitalo et al., 2015; Chakraborty et al., 2018; Maffeo and Aksimentiev, 2020). Bottom-up approaches have been most successfully used to study fluctuations within the duplex state, where the atomistic models on which they are built are best parameterised. Top-down approaches, by contrast, have found their application in the analysis of processes that involve DNA outside of its canonical B-form, including duplex hybridization (Ouldridge et al., 2013b), strand displacement (Srinivas et al., 2013; Irmisch et al., 2020),

stress-induced structural transitions (Romano et al., 2013; Wang and Pettitt, 2014; Sutthibutpong et al., 2016) and the properties of nanostructures with branched helices and single-stranded sections (Rovigatti et al., 2014; Engel et al., 2020).

Although highly-simplified, all of the coarse-grained models cited above attempt to represent the discrete, three-dimensional structure of DNA explicitly. An important role in our understanding of DNA is played by even simpler models. In thermodynamic terms, two classes of model have received particular attention. Firstly, the Peyrard-Bishop-Dauxois model and its variants have been used to probe the statistical properties of the duplex denaturation transition in the thermodynamic limit (Dauxois et al., 1993; Cocco and Monasson, 1999; Nisoli and Bishop, 2011). These models represent DNA through two or three continuous degrees of freedom per base pair.

A second approach dispenses with continuous degrees of freedom altogether, taking an Ising-like approach in which base pairs are either present or absent. Originally introduced by Poland and Scheraga to probe the duplex denaturation phase transition (Poland and Scheraga, 1966), the approach was adapted and carefully parameterized (SantaLucia, 1998; SantaLucia and Hicks, 2004; Huguet et al., 2010; Bae et al., 2020) to describe binding equilibria for strands of moderate length (oligonucleotides). It is difficult to overstate just how influential the nearest neighbour model has been, particularly in the development of nucleic acid nanotechnology, as it allows rational design of an ensemble of strands to produce the desired thermodynamics. The NUPACK software suite automates this process of system analysis and thermodynamics-based design by implementing the nearest-neighbour model (SantaLucia and Hicks, 2004). A number of attempts have been made to augment this thermodynamic model with realistic kinetics (Flamm et al., 2000; Xayaphoummine et al., 2005; Srinivas et al., 2013; Schaeffer et al., 2015).

At its simplest, the nearest-neighbour model allows a two-state approximation to the binding of A and B, in which the strands are either fully bound or fully dissociated. In this limit, the concentration of the product $[AB]$ can be estimated using the equation

$$\frac{[AB]}{[A][B]} = \exp\left(-\Delta H_{AB} - T\Delta S_{AB}/kT\right). \quad (1)$$

Here ΔH_{AB} and ΔS_{AB} are computed by summing contributions from each nearest-neighbor set of two base pairs, together with terms for helix initiation and various structural features, all of which are assumed to be temperature independent.

Another class of models ignores thermodynamics entirely, instead providing a continuum-level description of DNA mechanics. Most notably, DNA is frequently modelled as a semi-flexible polymer (or worm-like chain, WLC) characterised by a bending modulus (Kratky and Porod, 1949). This model can be augmented with an extensional modulus (Odijk, 1995) and a representation of twist with associated twist modulus (Yamakawa, 1977). It is also possible to consider coupling between the modes of deformation (Gore et al., 2006; Nomidis et al., 2019). As with the nearest-

neighbour model of DNA, the influence of these approaches is enormous, particularly within the biophysics community. The elastic rod is the starting point for understanding the geometry of DNA, and the null model against which results are compared and interpreted.

There is actually quite a large gap in complexity between mesoscopic models such as oxDNA and the continuum WLC models or the nearest neighbour model of thermodynamics. The time required to analyse the same system with these methods differs by many orders of magnitude. It is intriguing that, to our knowledge, there are few approaches that come close to bridging this gap. Fundamentally, it is not easy to combine the mechanics of semiflexible DNA as captured by the WLC, the geometry and topology of DNA structures, and the thermodynamics of DNA duplex formation as described in the nearest-neighbour model, in a representation that is simultaneously quantitatively useful and substantially simpler than the existing mesoscale models. Approaches such as Benham's description of melting in circularly negatively-supercoiled DNA (Fye and Benham, 1999) achieve this marriage in specific contexts. The variety of possible behaviour, however, and the sensitive interplay of topology, structure, mechanics and thermodynamics in many systems of interest, make the development of such models extremely hard and currently necessitate the application of coarse-grained models such as oxDNA.

In the rest of this pedagogical review, we first provide a high-level description of the basics of the oxDNA model and simulation techniques. We then present prototypical simulations to demonstrate key properties of oxDNA, and discuss how results from these simulations can be interpreted in terms of simpler DNA models at different length scales. While doing so, we discuss specific challenges in obtaining meaningful data from oxDNA simulations, and discuss where oxDNA provides added value. Initialisation files, processing scripts and supporting instructions are provided for all simulations presented here at (Sengar, 2021). This review should then serve as an introductory tutorial to applying oxDNA.

One drawback of this format is that the examples are presented as a fait accompli; just re-running the code will provide a limited experience of the real process of simulating oxDNA. We strongly encourage readers using this document as a tutorial to attempt to construct as much as possible of the simulations for themselves, and then to compare to the results obtained here. Alternatively, users may try to construct variants to simulate similar systems. Additional guidance on the nuts and bolts of running simulations can be found at Ref (oxDNA wiki, 2015; LAMMPS Documentation, 2021), where instructions on visualizing the output can also be found. In general, we have found that checking one or two snapshots of a simulation can avoid many wasted hours simulating and studying faulty systems.

3 THE OXDNA MODEL

The oxDNA model was originally developed to study the self-assembly, structure and mechanical properties of DNA

nanostructures, and the action of DNA nanodevices—although it has since been applied more broadly. To describe such systems, a model needs to capture the structural, mechanical and thermodynamic properties of single-stranded DNA, double-stranded DNA, and the transition between the two states. It must also be feasible to simulate large enough systems for long enough to sample the key phenomena. As discussed in **Section 2**, mesoscopic models in which multiple atoms are represented by a single interaction site are the appropriate resolution for these goals.

We will now outline the key features of the oxDNA model, the specific mesoscale model that is the focus of this review. While doing so, we note that there are effectively three versions of the oxDNA potential that are publicly available. The original model, oxDNA1.0 (Ouldridge et al., 2011), lacks sequence-specific interaction strengths, electrostatic effects and major/minor grooving. oxDNA1.5 adds sequence-dependent interaction strengths to oxDNA1.0 (Šulc et al., 2012), and oxDNA2.0 (Snodin et al., 2015) also includes a more accurate structural model, alongside an explicit term in the potential for screened electrostatic interactions between negatively charged sites on the nucleic acid backbone. In addition to these three versions of the DNA model, an RNA parameterisation "oxRNA" has also been introduced (Šulc et al., 2014).

It is worth noting that the three versions of oxDNA are very similar; most of the changes involve small adjustments of the geometry and strength of interactions. Structurally, the most significant change is the addition of a screened electrostatic interaction in oxDNA 2.0, which is typically small unless low salt concentrations are used. Moreover, subsequent versions of the model have been explicitly designed to preserve aspects of earlier versions that performed well. So oxDNA 1.5 and oxDNA 1.0 are very similar, except that oxDNA predicts sequence-dependent thermodynamic effects that are absent in oxDNA 1.0. oxDNA 2.0 is designed to preserve the thermodynamic and mechanical properties of oxDNA 1.5 at high salt as far as possible, but improves the structural description of duplexes (and structures built from duplexes) and allows for accurate thermodynamics at lower salt concentrations.

As a result, therefore, the discussion provided here for simulation of one version of the model largely applies to all. Moreover, it is worth noting that, even given the improved accuracy of oxDNA 2.0 in certain contexts, simulations of earlier versions of the model are still potentially valuable. oxDNA 2.0 comes into its own when it is essential to incorporate longer-range electrostatics at low salt concentrations, or when the detailed geometry of the helices are particularly important. A good example would be when simulating densely packed helices connected by crossover junctions in DNA origami (see **Section 8**). In other contexts, the reduced complexity of oxDNA 1.5 and hence its improved computational efficiency (along with slightly greater focus on basic thermodynamics and mechanics) may be beneficial. Further, it is often helpful to use oxDNA 1.0 in these contexts, as the comparison of versions 1.0 and 1.5 can help to distinguish sequence-dependent and generic effects.

In all three parameterisations, oxDNA represents each nucleotide as a rigid body with several interaction sites, namely the backbone, base repulsion, stacking and hydrogen-bonding sites, as shown in **Figure 1**. In oxDNA1.0 and oxDNA1.5, these sites are co-linear; the more realisic geometry of oxDNA2.0 offsets the backbone to allow for major and minor grooving.

Interactions between nucleotides depend on the orientation of the nucleotides as a whole, rather than just the position of the interaction sites. In particular, there is a vector that is perpendicular to the notional plane of the base, and a vector that indicates the direction of the hydrogen bonding interface. These vectors are used to modulate the orientational dependence of the interactions, which allows the model to represent the coplanar base stacking, the linearity of hydrogen bonding and the edge-to-edge character of the Watson–Crick base pairing. Furthermore, this representation allows the encoding of more detailed structural features of DNA, for example, the right-handed character of the double helix and the anti-parallel nature of the strands in the helix.
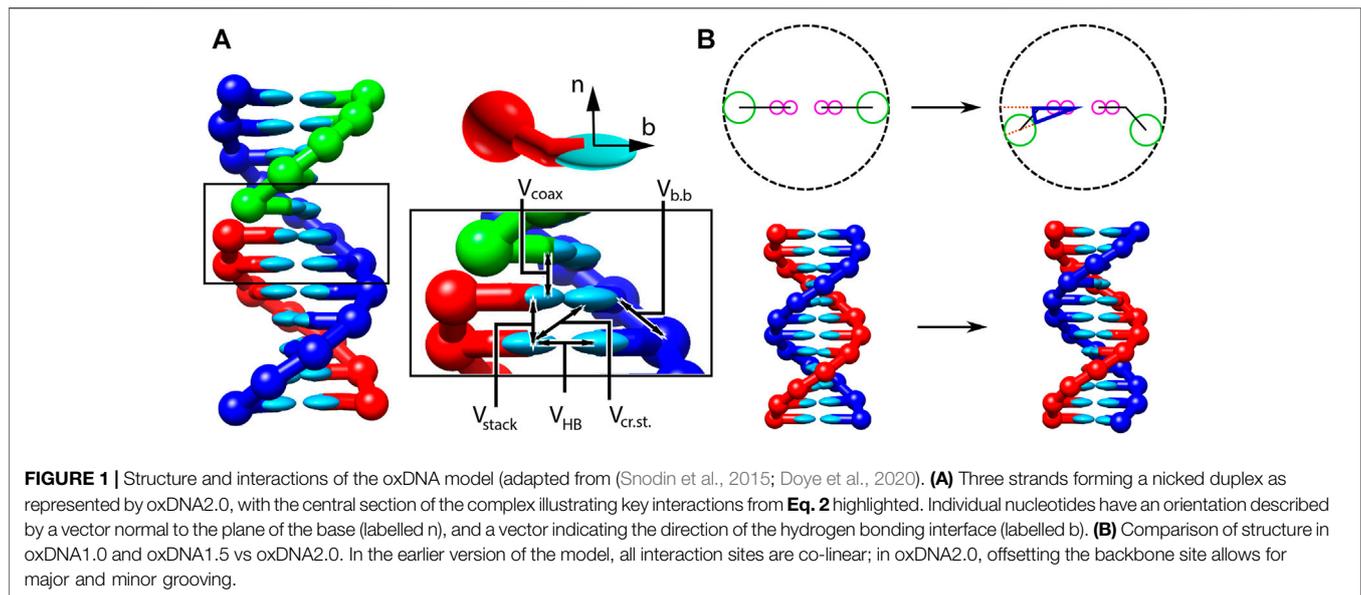
The potential energy of the system is calculated as:

$$V_0 = \sum_{\langle ij \rangle} \left(V_{b.b.} + V_{stack} + V'_{exc} \right) + \sum_{i,j \notin \langle ij \rangle} \left(V_{HB} + V_{cr.st.} + V_{exc} + V_{coax} \right),$$

(2)

with an additional screened electrostatic repulsion term for oxDNA 2.0. In **Eq. 2**, the first sum is taken over all pairs of nucleotides that are nearest neighbors on the same strand and the second sum comprises all remaining pairs. The terms represent backbone connectivity ($V_{b.b.}$), 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 ($V_{cr.st.}$) across the duplex axis and coaxial stacking (V_{coax}) across a nicked backbone. The excluded volume and backbone interactions are a function of the distance between repulsion sites. The backbone potential is a spring potential mimicking the covalent bonds along the strand. All other interactions depend on the relative orientations of the nucleotides and the distance between the hydrogen-bonding and stacking interaction sites.

A crucial feature of the oxDNA model is that the double helical structure is driven by the interplay between the hydrogen-bonding, stacking and backbone connectivity bonds. The stacking interaction tends to encourage the nucleotides to form co-planar stacks; the fact that this stacking distance is shorter than the backbone bond length results in a tendency to form helical stacked structures. In the single-stranded state, these stacks can easily break, allowing the single strands to be flexible. The geometry of base pairing with a complementary strand locks the nucleotides into a much more stable double helical structure.

The model was deliberately constructed with all interactions pairwise (i.e., only involving two nucleotides, which are taken as rigid bodies). This pairwise character allows us to make effective use of cluster-move Monte Carlo (MC) algorithms, which provide efficient equilibrium sampling (see **Section 8**).

FIGURE 1 | Structure and interactions of the oxDNA model (adapted from (Snodin et al., 2015; Doye et al., 2020). **(A)** Three strands forming a nicked duplex as represented by oxDNA2.0, with the central section of the complex illustrating key interactions from **Eq. 2** highlighted. Individual nucleotides have an orientation described by a vector normal to the plane of the base (labelled n), and a vector indicating the direction of the hydrogen bonding interface (labelled b). **(B)** Comparison of structure in oxDNA1.0 and oxDNA1.5 vs oxDNA2.0. In the earlier version of the model, all interaction sites are co-linear; in oxDNA2.0, offsetting the backbone site allows for major and minor grooving.

It is convenient to use reduced units to describe lengths, energies and times in the system. A summary of the conversion of these "oxDNA units" to SI units is provided in **Supplementary Appendix A**.

4 SIMULATING THE MODEL: MOLECULAR DYNAMICS (MD) VS VIRTUAL-MOVE MONTE CARLO (VMMC)

The oxDNA model is far too complicated to approach analytically. Publicly released code to simulate oxDNA is available as a standalone package (oxDNA wiki, 2015), or as a module (LAMMPS Documentation, 2021) for the popular LAMMPS simulation software. We note that in the process of preparing this manuscript, a small error was identified in the potential as implemented in LAMMPS (see **Supplementary Appendix B**). This error is only really noticeable when simulating unpaired DNA bases. Nonetheless, we recommend that potential users of the LAMMPS implementation wait for the stable LAMMPS release in Summer 2021. The fix will be verified through an erratum attached to the original publication Henrich et al. (2018).
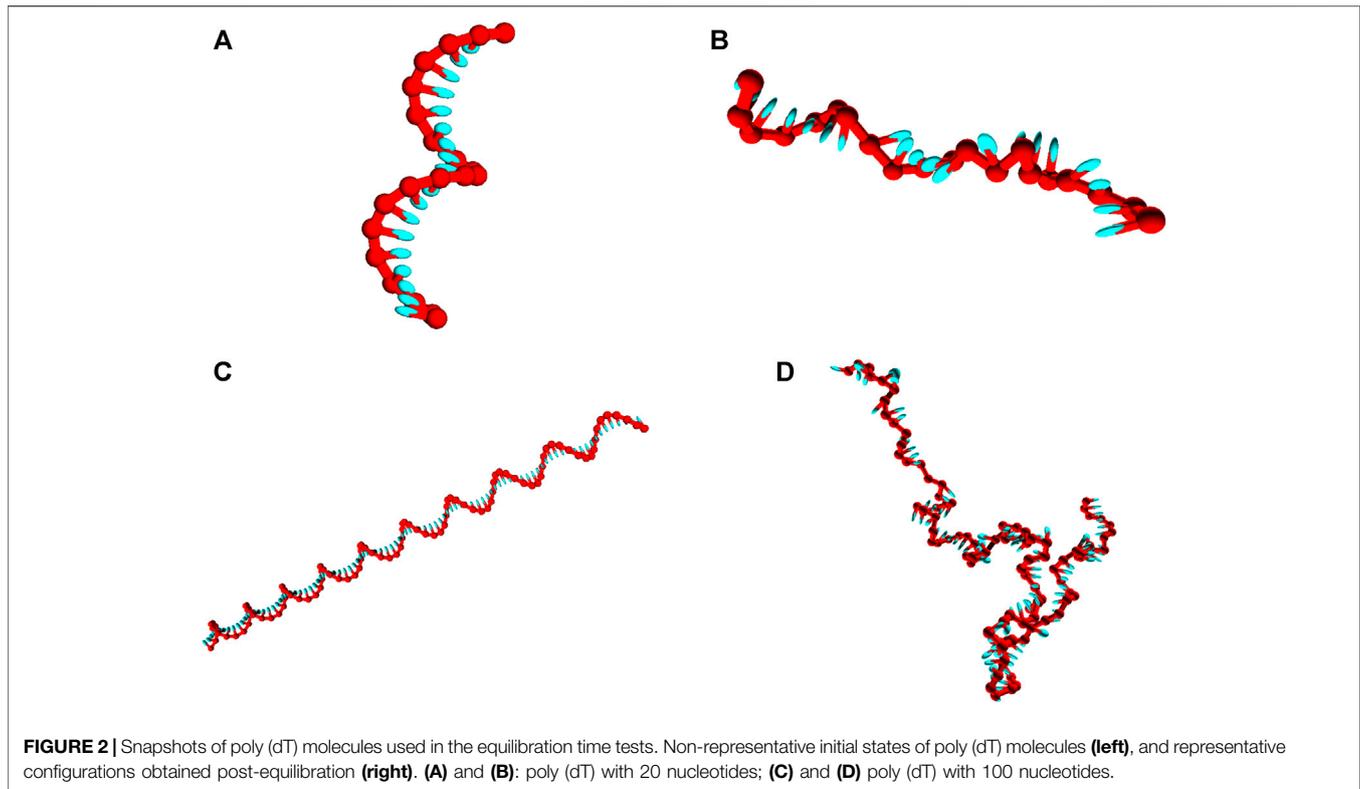
There are two broad types of simulation technique that can be applied to probe the model: molecular dynamics (MD) and Monte Carlo (MC). Molecular dynamics (Frenkel and Smit, 2002) algorithms evolve their constituent molecules according to Newton's laws of motion, and so are a natural choice for simulating particle systems. For coarse-grained models such as oxDNA, in which the solvent is implicit, it is necessary to include a thermostat to both set the temperature and ensure diffusive rather than ballistic dynamics. The default MD algorithm for the standalone version of oxDNA is an Andersen-like algorithm (Russo et al., 2009), in which particle velocities and angular velocities are resampled from a Boltzmann distribution with a frequency that sets the effective diffusion coefficient. In the LAMMPS implementation, the model utilises a Langevin thermostat for rigid bodies (Davidchack et al., 2015), which applies small friction- and noise-based updates to the momentum and angular momentum at each step. The relative size of these contributions sets the temperature.

A challenge of MD simulations is that when strong, short-ranged interactions are present—as in oxDNA—they place a limit on the maximum integration time step that can be used while preserving numerical stability. Interestingly both the Andersen-like and Langevin thermostats act to stabilise the simulations, allowing larger time steps to be used than if the equations of motion were integrated without noise or drag to generate energy-conserving, ballistic motion.

Both MD algorithms generate dynamical trajectories that can be used to probe system kinetics (more on this in **Section 7**). However, it is also common to use MD to take equilibrium averages over the configurations of a particular system. In the limit of small time steps, both Andersen-like and Langevin algorithms will converge on a steady state in which they sample configurations $\mathbf{x}$ from the Boltzmann distribution $p_{eq}(\mathbf{x}) \propto \exp(-\beta V_0(\mathbf{x})/k_B T)$, where $V_0(\mathbf{x})$ is the potential energy of the model. How small the step size needs to be depends on a number of details, such as the strength of coupling to the thermostat. For parameters that have become an unofficial default for oxDNA, we illustrate the accuracy of the algorithm as a function of step size in **Supplementary Appendix B**.

Monte Carlo (MC) (Metropolis and Ulam, 1949) simulations are an alternative approach for sampling from the same Boltzmann distribution, but evading the drawbacks caused by the presence of a timestep altogether. In the standard MC approach (Frenkel and Smit, 2002), configurational moves $\mathbf{x} \rightarrow \mathbf{y}$ are proposed randomly, with a symmetric probability distribution that satisfies $p_{gen}(\mathbf{x} \rightarrow \mathbf{y}|\mathbf{x}) = p_{gen}(\mathbf{y} \rightarrow \mathbf{x}|\mathbf{y})$. If these proposed moves are accepted with a probability

FIGURE 2 | Snapshots of poly (dT) molecules used in the equilibration time tests. Non-representative initial states of poly (dT) molecules **(left)**, and representative configurations obtained post-equilibration **(right)**. **(A)** and **(B)**: poly (dT) with 20 nucleotides; **(C)** and **(D)** poly (dT) with 100 nucleotides.

$$p_{\text{acc}}(\mathbf{x} \to \mathbf{y}) = \max(\exp(-(V_0(\mathbf{y}) - V_0(\mathbf{x}))/k_B T, 1)),$$ then the Boltzmann distribution $p_{\text{eq}}(\mathbf{x}) \propto \exp(-\beta V_0(\mathbf{x}))$ is the stationary distribution of the simulation and a long simulation will sample from that distribution, assuming ergodicity.

In principle, the moves $\mathbf{x} \to \mathbf{y}$ can be arbitrarily large without leading to errors, since it is not necessary to integrate the derivative of the potential, only calculate its values at the endpoints. However, standard MC techniques incorporate sequential updates of individual particles as the moves $\mathbf{x} \to \mathbf{y}$. For a model of a strongly-attractive system such as oxDNA, these moves must be extremely small or the acceptance factor will always be small. The result is painfully slow equilibration, particularly if large scale movements of strands is required to observe it.

Virtual-move Monte Carlo (VMMC) (Whitelam and Geissler, 2007; Whitelam et al., 2009) is an alternative that circumvents the drawbacks of MC algorithms. VMMC first proposes a single particle move, then generates a co-moving cluster of particles based on which interactions are best preserved by moving the particles in unison while ensuring that the correct, detailed-balanced stationary distribution is retained. The cluster building process is based on assessing the change in pairwise interactions, and so VMMC is especially suited to oxDNA, which has exclusively pairwise interactions. We have implemented the variant from the appendix of Ref (Whitelam et al., 2009). in the standalone code.

For those with limited experience of simulating oxDNA, it is not obvious whether VMMC or MD is the optimal approach to sampling a given system. We illustrate the relative efficiencies of

the two algorithms when simulating ssDNA of length 20, 100 and 1,000 bases, in terms of the computational time required to reach states representative of equilibrium from the same unrepresentative starting condition.

We simulate the poly (dT) molecules with (a) 20 bases, (b) 100 bases and (c) 1,000 bases, using oxDNA1.0 (Ouldridge et al., 2011). Simulations are performed at $T = 27^{\circ}\text{C}$, in a periodic box of 20, 100 and 1,000 simulation units for 20, 100 and 1,000 bases, respectively. For the MD simulations, we simulate for 60,000 simulation units of time (a nominal 182 ns), with a time step of $dt = 0.003$ (see **Supplementary Appendix B**); each simulation therefore has 2×10^7 steps in total. For VMMC, we attempted 60,000 VMMC steps per particle. The proposed moves are: rotation about a random axis, through an angle up to 0.22 radians; and translation through a distance of up to 0.22 units. These choices produce a nice balance of cluster sizes, ranging from individual nucleotides to entire strands.

Strands are initialized in a fully stacked, helical conformation as illustrated in **Figures 2A,C**, and relax to more-representative, partially-stacked conformations (**Figures 2B,D**) as the simulation is run. The relaxation of the strands is associated with an increase in the potential V_0, and so we illustrate equilibration by plotting that potential averaged over 20 independent simulations as a function of simulation progress in **Figure 3**. The average value of the potential in equilibrium, $\langle V_0 \rangle_{\text{eq}}$, can be approximated by the average over the data collected in the second half of the simulations. We then estimate the equilibration time scale as the time required for $\langle V_0(t) \rangle - \langle V_0 \rangle_{\text{eq}}$ to reach $1/e$ of its initial value for the first time.
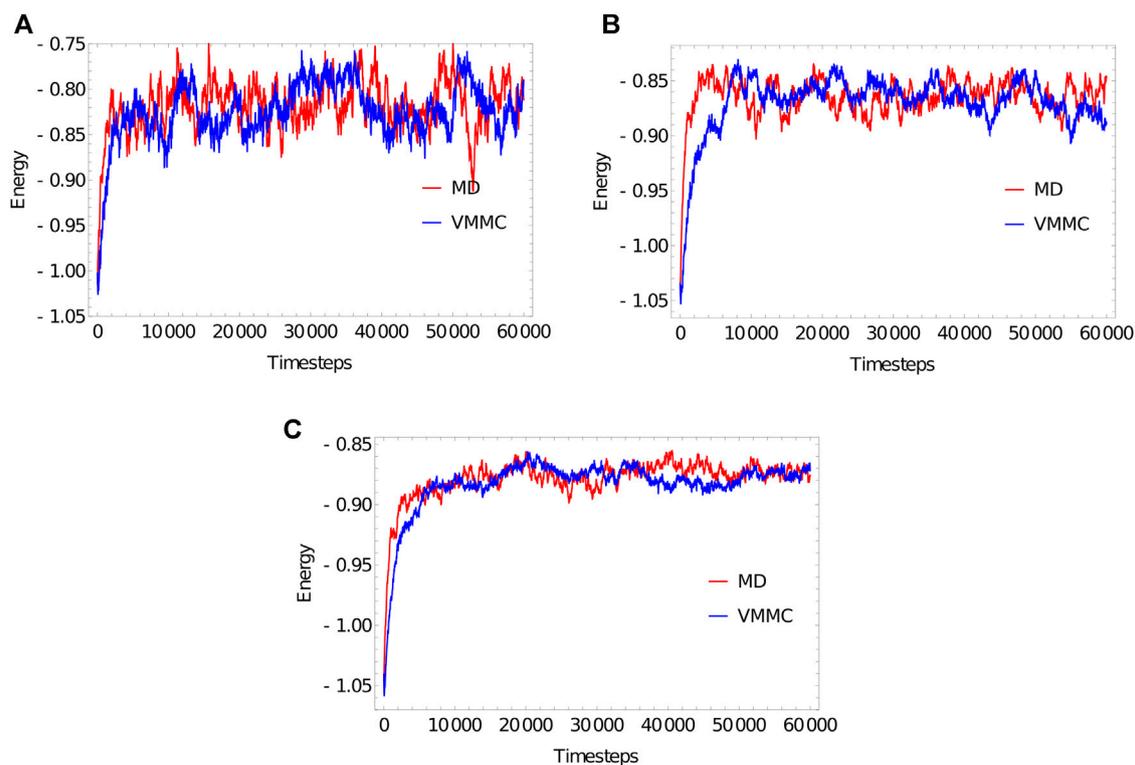
FIGURE 3 | Equilibration plots for the ploy (dT) molecules with **(A)** 20 bases, **(B)** 100 bases, **(C)** 1,000 bases, obtained as averages over 20 independent simulations. For both MD and VMMC, the potential V_0 is plotted as a function of simulation progress, in units of reduced time (3.03 ps) for MD and attempted steps per particle for VMMC.

TABLE 1 | Computational time for equilibration of poly (dT) molecules of various lengths. Simulations were performed using a single core Intel(R) Core(TM) i5-4300U CPU @ 1.90GHz.

Strand length	Total runtime (MD) in seconds	Total runtime (VMMC) in seconds	Equilibration time as a fraction of total runtime (MD)	Equilibration time as a fraction of total runtime (VMMC)
20	409.67	14.804	0.0118	0.0244
100	1804.18	147.57	0.0128	0.0287
1,000	32493.0	4187.03	0.0134	0.0299

The "simulation progress" axes in **Figure 3** are not directly comparable for MD and VMMC; one measures simulation time, the other attempted VMMC steps per particle. The most relevant quantity is the actual computational time required to equilibrate the system on a given architecture; in simple contexts, this time is also indicative of the speed with which the algorithm samples the equilibrium ensemble. **Table 1** shows the total runtime of the simulations and the equilibration time as a fraction of that runtime. In computational time, the VMMC algorithm is able to equilibrate the poly (dT) molecules more quickly (compared to MD algorithm) in all the three cases. VMMC is around 15 times as fast for the 20-nucleotide strand, dropping to around 4 times as fast for the 1000-nucleotide strand. The large moves available to VMMC, and the lack of a requirement to differentiate potentials, provide this benefit. Note, however, that the ratio of the

equilibration times for MD to VMMC algorithms decreases as the system size increases; as can be seen in **Table 1**, the equilibration time for VMMC simulations increases super-linearly with system size. This super-linear increase arises because more steps must be taken to equilibrate larger systems *and* because larger clusters tend to be built for larger systems, resulting in each step taking longer on average.

The relative efficiency of VMMC and MD approaches will depend to some degree on the choice of damping parameters and seed moves; we have not carefully optimised our choices for either technique, but have used values that generally work well. The relative efficiency will also depend on the particular system: VMMC lends itself to systems in which large movements are important. Our metric for efficiency (speed with which systems reach a state that is representative of equilibrium) is also fairly

crude. Nonetheless, the general rule of thumb that VMMC is more efficient for smaller systems—particularly those with significantly fewer than 1,000 nucleotides, such as the sampling of duplex formation in oligonucleotides - is a helpful one. It is also particularly easy to enhance VMMC using umbrella sampling, as explained in **Section 6.1.1**.

For sufficiently large systems, such as DNA origami, MD should equilibrate faster, and therefore provide improved sampling. Another major advantage of the MD approach is much more facile parallelisation when simulating large systems. The standalone code allows for parallel simulation on GPUs (graphical processing units), and the LAMMPS module for parallel simulations across multiple CPUs (central processing units) using MPI. These approaches are demonstrated in **Section 8**.

5 MECHANICAL PROPERTIES OF DNA

The mechanical properties of DNA are central to its role across nanotechnological, biophysical and biological contexts. DNA's flexibility and response to applied stress determine the conformation and accessibility of the genome inside cells (Lewis et al., 1996; Nikolov et al., 1996; Widom, 2001; Richmond and Davey, 2003). Moreover, not only is the stiffness of dsDNA important in maintaining the conformation of DNA nanostructures, but the relative flexibility of ssDNA crucially allows for joints and flexible hinges. These properties are widely-studied in bulk and single-molecule experiments *in vitro* (Crothers et al., 1992; Smith et al., 1996; Strick et al., 1996; Wang et al., 1997; Rivetti et al., 1998; Mills et al., 1999; Podtelezhnikov et al., 2000; Dessinges et al., 2002; Bryant et al., 2003; Seol et al., 2004; Fujimoto et al., 2006; Gore et al., 2006; Lionnet et al., 2006; Seol et al., 2007; Du et al., 2008; Forth et al., 2008; Mosconi et al., 2009; Demurtas et al., 2009; Brutzer et al., 2010; Gross et al., 2011; Salerno et al., 2012; Tempestini et al., 2013; Fields et al., 2013; Le and Kim, 2014; Kim et al., 2015).

It is therefore essential that a coarse-grained model provides a reasonable representation of these properties. In this section, we both discuss the mechanical properties of oxDNA, and show how to construct simulations that can probe these properties.

5.1 Stiffness of Duplex and Single-Stranded DNA

The most common metric used to quantify the stiffness of DNA is the persistence length, defined in the textbook of Cantor and Schimmel as (Cantor and Schimmel, 1980)

$$L_{ps} = \frac{\langle \mathbf{L} \cdot \mathbf{l}_0 \rangle}{\langle l_0 \rangle}. \tag{3}$$

Here, $\mathbf{L}$ is the end-to-end vector of the polymer and $\mathbf{l}_0$ represents the vector between the first two monomer units. dsDNA is most commonly thought of as a semi-flexible polymer or wormlike chain (Kratky and Porod, 1949). In this
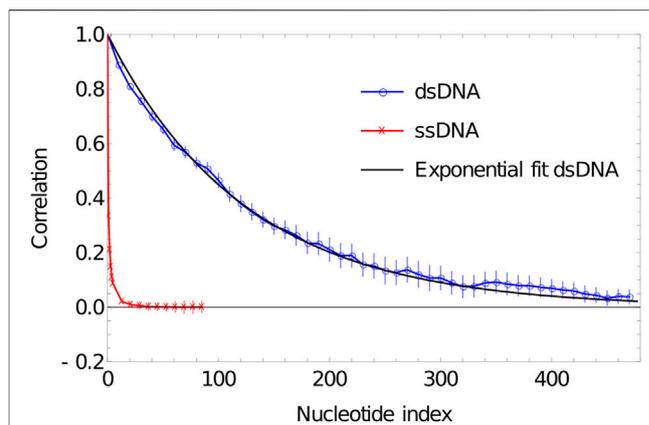
FIGURE 4 | Plot of correlation of inter-segment vectors vs distance (number of base pairs along the DNA) for 500 dsDNA (blue curve) and 100 ssDNA (red curve). dsDNA fitted with an exponential decay with a decay constant of 0.0076134.

picture, the discrete series of inter-base pair vectors are approximated as a continuous, differentiable polymer axis with a quadratic free energy of curvature. For an infinitely long, semi-flexible polymer, correlations in the alignment of the polymer axis decay exponentially with separation, with a decay rate given determined by L_{ps}. When translated back to the language of inter-base-pair vectors, we obtain

$$\frac{\langle \mathbf{l}_n \cdot \mathbf{l}_0 \rangle}{\langle l_0 \rangle^2} = \exp\left(- n\langle l_0 \rangle \big/ L_{ps}\right), \tag{4}$$

where $\mathbf{l}_n$ is the vector between base pair $n - 1$ and base pair n.

It is relatively straightforward to both assess whether the wormlike chain model is a good model for oxDNA, and to extract L_{ps}. We simply simulate a duplex system for long enough to sample a representative set of configurations, calculate the correlation between inter-base-pair vectors as a function of separation, and fit the results to the exponential decay of **Eq. 4**.

In **Figure 4**, we plot the results of such a procedure. To obtain these data, we simulate a DNA duplex of length 500 base pairs at 27°C using oxDNA1.5. We perform 20 VMMC simulations with 2×10^7 attempted steps per particle for VMMC. The first 10^5 moves are treated as an initialization period and no data is collected. Additionally, the base pairs at the ends of the duplex are more flexible than those well within the bulk; to obtain properties representative of bulk DNA, we therefore do not include the five base pairs at either end in our analysis. Correlations are calculated from the default configurational outputs of the model, using the code provided in (Sengar, 2021). In order to obtain a good sample, it is helpful to output these configurations with a high frequency; we use a small value for the parameter to output energy configurations after a single VMMC move per particle.

As is evident from **Figure 5**, the correlation of the duplex axis indeed follows an exponential fall-off, to within sampling error. Fitting **Eq. 4** to $\ln\left(\frac{\langle \mathbf{l}_n \mathbf{l}_0 \rangle}{\langle l_0 \rangle^2}\right)$ gives $L_{ps} \approx 131\langle l_0 \rangle = 131 \times$
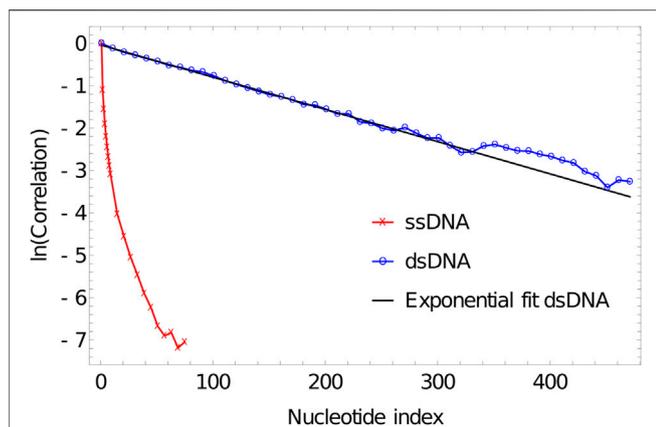
FIGURE 5 | Log plot of correlation of inter-segment vectors vs distance (number of base pairs along the DNA) for 500 dsDNA (blue curve) and 100ssDNA (red curve). dsDNA fitted with an exponential decay with a decay constant of 0.0076134.

$0.4118 \approx 53.95$ simulation units= 45.91 nm, consistent with experimental estimates of 40–50 nm (120–150 base pairs) at high [Na+] concentrations (Savelyev, 2012; Herrero-Galán et al., 2013).

Indeed, more generally, the mechanical properties of double-stranded oxDNA are well-described by a semiflexible polymer model, and its torsional and extensional moduli have been analysed elsewhere (Ouldridge et al., 2011; Matek et al., 2015). Significant deviations from this behaviour - such as sharp kinks facilitated by broken base pairs—are generally only observed when large stresses are applied to the molecule (Romano et al., 2013; Matek et al., 2015), in agreement with experiment.

ssDNA behaves very differently in oxDNA. In **Figure 4**, we plot the correlations of backbone-site-to-backbone-site vector for a 100-base poly (dT) ssDNA, obtained from running simulation in oxDNA1.5 at 27°C. We perform 20 VMMC simulations with 2×10^7 attempted steps per particle. The first 10^5 moves are treated as an initialization period and no data is collected. For these simulations, we have set the stacking strength between the nucleotides to zero (the consequences of non-zero stacking strength will be addressed in Section 5.2). From **Figure 4** and **Figure 5**, it is apparent that the correlation drops very rapidly, meaning that unstacked ssDNA is very flexible in oxDNA, as it should be; adjacent backbone-to-backbone vectors can bend through a large angle. But importantly, it is worth noting that the drop in correlation between vectors with separation along the polymer cannot be well described by an exponential as in **Eq. 4**. The convexity of $\ln\left(\frac{\langle l_n \cdot l_0 \rangle}{\langle l_0 \rangle^2}\right)$ is indicative of more distant backbone-to-backbone vectors being aligned more strongly than would be expected from the alignment of two adjacent backbone-to-backbone vectors.

The reason for this behaviour is that it is the excluded volume of nucleotides that gives unstacked ssDNA its "stiffness" in oxDNA. The excluded volume of nucleotides discourages ssDNA from folding back on itself, but importantly it leads to very different polymer properties than assumed in common polymer models such as the freely-jointed chain and the wormlike chain. For these classic polymer models, the statistical properties are entirely determined by interactions between parts of the polymer that are adjacent along the backbone, whereas the curvature of $\ln\left(\frac{\langle l_n \cdot l_0 \rangle}{\langle l_0 \rangle^2}\right)$ in **Figure 5** is indicative of interactions between more distant points along the polymer contour playing a role.

As a result, using a wormlike chain with a given L_{ps} (or a freely jointed chain with a given Kuhn length) to understand ssDNA in oxDNA is misleading. The overall tendency of the polymer to swell to fill a large volume - due to its excluded volume - would suggest a far greater degree of local stiffness than actually present. This effect is retained even when stacking between adjacent nucleotides is included.
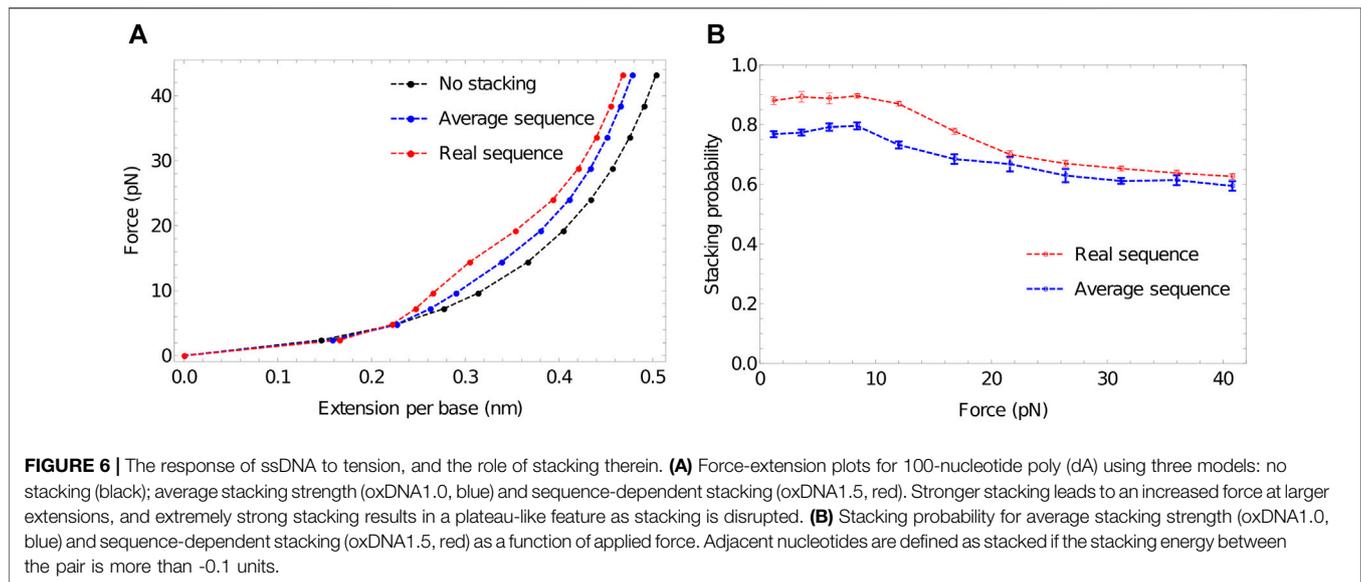
Importantly, these complexities also apply to physical ssDNA, as well as the oxDNA model. Single DNA have linear dimensions on the order of 1 nm, and experimental attempts to measure the mechanical properties of oxDNA (usually reported as persistence lengths) are of a similar order of magnitude (Smith et al., 1996; Rivetti et al., 1998; Mills et al., 1999; Murphy et al., 2004). Describing ssDNA in this way is not self-consistent; any polymer with this cross-section and flexibility would be strongly affected by excluded volume, so these models cannot be accurate. The result has been that experiments on large scale properties of relaxed ssDNA (Rivetti et al., 1998; Murphy et al., 2004), which are sensitive to excluded volume effects, tend to produce larger estimates for quantities like L_{ps} than experiments on shorter sections of ssDNA, or ssDNA under high tension (Smith et al., 1996; Rivetti et al., 1998).

Low salt concentrations, which lead to weaker screening of electrostatic interactions between non-adjacent nucleotides make the above effect stronger (Smith et al., 1996; Dessinges et al., 2002). Base-pairing interactions in non-homopolymeric ssDNA have a confounding effect; the formation of secondary structure tends to condense the strand, making it appear more flexible when its statistics are modelled with a wormlike chain or a freely-jointed chain (Smith et al., 1996). Overall, as for oxDNA, simple descriptions of the mechanical properties of physical ssDNA should be treated with caution.

5.2 Response of ssDNA to Tension

A common mechanism for probing the mechanical properties of DNA is to apply force, whether torsional (Strick et al., 1996; Bryant et al., 2003; Forth et al., 2008; Mosconi et al., 2009; Brutzer et al., 2010; Salerno et al., 2012; Tempestini et al., 2013), extensional (Smith et al., 1996; Strick et al., 1996; Wang et al., 1997; Dessinges et al., 2002; Seol et al., 2004; Gore et al., 2006; Seol et al., 2007; Huguet et al., 2010; Gross et al., 2011) or shearing (Forth et al., 2008; Hatch et al., 2008; Mosconi et al., 2009; van Mameren et al., 2009; Brutzer et al., 2010; Salerno et al., 2012; Tempestini et al., 2013; Wang and Ha, 2013).

Experiments have focused on both the elastic properties associated with small deformations (Smith et al., 1996; Strick et al., 1996; Wang et al., 1997; Bryant et al., 2003), and large scale structural transitions (Smith et al., 1996; Strick et al., 1996; Dessinges et al., 2002; Bryant et al., 2003; Seol et al., 2004; Gore et al., 2006; Seol et al., 2007; Hatch et al., 2008; Forth et al., 2008; Mosconi et al., 2009; van Mameren et al., 2009;

FIGURE 6 | The response of ssDNA to tension, and the role of stacking therein. **(A)** Force-extension plots for 100-nucleotide poly (dA) using three models: no stacking (black); average stacking strength (oxDNA1.0, blue) and sequence-dependent stacking (oxDNA1.5, red). Stronger stacking leads to an increased force at larger extensions, and extremely strong stacking results in a plateau-like feature as stacking is disrupted. **(B)** Stacking probability for average stacking strength (oxDNA1.0, blue) and sequence-dependent stacking (oxDNA1.5, red) as a function of applied force. Adjacent nucleotides are defined as stacked if the stacking energy between the pair is more than -0.1 units.

Brutzer et al., 2010; Gross et al., 2011; Salerno et al., 2012; Wang and Ha, 2013; Tempestini et al., 2013).

Applying external tension is relatively straightforward in molecular simulation; there are more subtleties associated with applying boundary conditions for external torsion (Matek et al., 2012; Matek et al., 2015), but it is also possible. In the case of oxDNA, small external stresses have been used to help parameterise and characterise the model (Ouldridge et al., 2011; Matek et al., 2015; Skoruppa et al., 2017; Nomidis et al., 2019); larger stresses have been applied to provide insight into experiments on structural transitions (Matek et al., 2012; Romano et al., 2013; Wang and Pettitt, 2014; Matek et al., 2015; Mosayebi et al., 2015; Wang et al., 2015; Engel et al., 2018; Desai et al., 2020).

Systems with internally-induced stress, where the drive to form base pairs in one part of an assembly applies stress to another part, have also been studied (Harrison et al., 2015; Sutthibutpong et al., 2016; Wang and Pettitt, 2016; Wang et al., 2017; Tee and Wang, 2018; Caraglio et al., 2019; Harrison et al., 2019; Engel et al., 2020; Fosado et al., 2021; Park et al., 2021).

As an example, in this section we demonstrate the force-extension properties of ssDNA as represented by oxDNA. Optical tweezer experiments with ssDNA have a long history (Smith et al., 1996). These original experiments with naturally-occurring DNA exhibited formation and stabilization of secondary structure in high salt conditions and low-moderate force, although this was not explicitly modelled at the time. The presence of this secondary structure makes simulation of DNA heteropolymers hard; it is challenging to equilibrate a long strand with many competing base-pairing configurations (we have had some success using methods based on parallel tempering (Romano et al., 2013). Instead, therefore, we simulate 100-nucleotide-long homopolymeric poly (dA) using oxDNA1.0 and oxDNA1.5.

The helicity in oxDNA is driven by stacking interactions between adjacent nucleotides. As is evident from **Figure 2**, this stacking has a residual effect on the structure of ssDNA

strands, which are partially stacked in equilibrium. We will use force-extension simulations to probe the consequences of single-stranded stacking in oxDNA.

For these simulations, which are similar to original results in (Šulc et al., 2012), we use both a version of the parameters with no stacking interacting, a sequence-averaged stacking interaction (oxDNA1.0), and a sequence-specific stacking interaction (oxDNA1.5) for which poly (dA) has the strongest interaction of all sequences. All simulations are performed at $T = 27°C$, in a periodic box of length 100 simulation units with each simulation running for 4×8 steps with $dt = 0.005$ which equals to a total run time of 2×10^6 units. Four sets of simulations are performed for 12 difference values of force.

Figure 6A shows that extensive stacking has only a moderate effect on the force-extension properties of the ssDNA at low force. In the sequence-dependent model, poly (dA) is close to 90% stacked at 27°C—see **Figure 6B**. However, the increased stiffness due to the tendency to form stacked single helices (akin to the initial state in **Figure 3**) is counteracting by the shorter end-to-end distance of the backbone when it is forced to wind around the helix.

At larger forces, however, we clearly see a signal of stronger stacking. Larger force is required to extend the strands with stronger stacking, and a plateau-like feature is evident in the system with the strongest stacking. A similar plateau was observed by Seol *et al.* for RNA stretching (poly(A) and poly(C)) (Seol et al., 2004; Seol et al., 2007) but was absent for poly(U). Those authors hypothesised that the plateau arises as the shorter end-to-end distance in helical stacked confirmations becomes prohibitive; additional force is then required to disrupt the stacking interaction to allow further extension, causing an increase in the gradient of the force-extension curve. After the bases have unstacked, the gradient becomes less steep again.

Broadly speaking, this explanation is borne out by oxDNA. Notably, however, Seol *et al.* (Seol et al., 2004; Seol et al., 2007)

concluded that a relatively low stacking probability should give a pronounced plateau. By contrast, in **Figure 6B**—obtained by probing configuration output files to assess the degree of stacking (Sengar, 2021) we see that strands with an initial stacking probability of 78% show only a hint of the plateau. This discrepancy arises because, in the minimal model of Seol *et al.* (Seol et al., 2004; Seol et al., 2007), even a single pair of stacked nucleotides has a much shorter end-to-end distance along the ssDNA backbone than an unstacked pair. The explicit representation of 3D structure in oxDNA, however, captures the fact that the shortening of the end-to-end distance along the DNA backbone is only significant when several bases in a row are stacked into a helix, so that the backbone really has to wrap back round upon itself. As a result, extension can occur while disrupting only a fraction of the stacking interactions, and ssDNA in oxDNA remains significantly stacked even at high force (**Figure 6B**).

While oxDNA's representation of the polynucleotide backbone is simplistic, these geometrical arguments also apply to physical DNA - suggesting that even weak plateau-like behaviour in ssDNA force-extension curves is evidence of strong stacking, and the absence of a plateau is not proof of an absence of stacking. More generally, this system is indicative of the value that oxDNA can provide. The system involves an interplay between basic structure, mechanics and thermodynamics of ssDNA. When applied, oxDNA reveals subtleties that are not directly apparent from a more minimal model. Indeed, it is quite common to construct very simple models to interpret biophysical experiments on the mechanical properties of DNA (Hatch et al., 2008; Qu and Zocchi, 2011; Salerno et al., 2012; Vafabakhsh and Ha, 2012; Fields et al., 2013; Tempestini et al., 2013; Wang and Ha, 2013; Meng et al., 2014); simulations with oxDNA often reveal physically reasonable relaxation mechanisms that aren't factored into these simpler models (Harrison et al., 2015; Matek et al., 2015; Mosayebi et al., 2015; Skoruppa et al., 2017; Harrison et al., 2019). At this stage, it is also worth highlighting a general virtue of coarse-grained models that is apparent in these simulations. It is very simple just to switch off interactions—such as the stacking here—to isolate the effect those interactions have on the system. Doing so can be incredibly helpful in interpreting the physical cause of experimental signals.

6 THERMODYNAMIC SIMULATIONS WITH OXDNA

6.1 Duplex Formation Thermodynamics

As well as representing the structure and mechanical properties of ssDNA and dsDNA, oxDNA is also designed to capture the thermodynamics of the hybridization transition from ssDNA to dsDNA. Needless to say, accurately capturing the thermodynamics of this transition is essential for any model hoping to describe biological and nanotechnological processes involving the forming and disruption of base pairs.

To assess the thermodynamics of a simple duplex, it is typical to simulate an isolated pair of strands in a periodic cell that is large enough to prohibit self-interactions (unit cell size of $>2n$ oxDNA length units, where n is the duplex length, is generally sufficient). Given a sufficiently long VMMC or MD simulation, the fraction of time spent in the bound state can be estimated and used to infer quantities such as melting temperatures, as outlined below.

However, particularly for longer strands, simulating this process can be prohibitively slow. For two short strands in solution, the vast majority of configurations have well-separated strands and no base-pairing interactions. Enthalpically favourable base-pairing provides a compensatory advantage to configurations with many well-formed base pairs (fully-formed duplexes). To obtain a good estimate of the fraction of strands bound in equilibrium, it is necessary to pass between these two sub-ensembles (completely unbound and fully bound) many times; as a rule of thumb, we have found that around 10 interconversions will start to provide meaningful statistics.

Unfortunately, interconversion requires the system to transition through states with only one or two base pairs that benefit neither from the large ensemble of configurations accessible to dissociated strands, nor the favourable interactions of fully-bound strands. These configurations with $Q = 1, 2$ base pairs are rare in the equilibrium ensemble, and have a relatively high free energy

$$F(Q) = -kT\ln\bigl(p_{\text{eq}}(Q)\bigr) + C, \tag{5}$$

where C is a Q-independent constant, and $p_{\text{eq}}(Q)$ is the probability of observing Q base pairs in equilibrium. The high free energy of these intermediate states makes dissociation and association rare event processes that are challenging to sample directly.
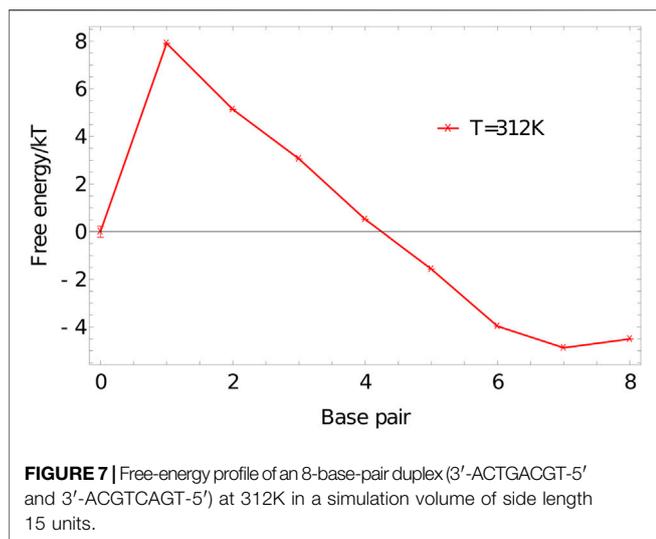
6.1.1 Umbrella Sampling

To overcome this difficulty, simulations can be augmented with umbrella sampling (Torrie and Valleau, 1977). For a system with coordinates x (in our case, nucleotide positions and orientations), umbrella sampling involves identifying a collective order parameter $\lambda(x)$ for the transition of interest, and then applying a bias $W(\lambda(x))$ to force the system to occupy otherwise undesirable values of $\lambda(x)$ that lie along the transition path more frequently. Unbiased statistical averages can be extracted from these biased samples using

$$\langle A(x)\rangle_{\text{eq}} = \left\langle \frac{A(x)}{W(\lambda(x))} \right\rangle_{\text{biased}}, \tag{6}$$

where $A(x)$ is a quantity of interest. Essentially, the contribution of each configuration sampled to the average is reduced by a factor of the bias applied.

A common approach with umbrella sampling is to perform a series of separate simulations with very strong biases tightly centred on distinct values of $\lambda(x)$. Simulations centred on adjacent values of $\lambda(x)$ can then be knitted together using procedures such as the Weighted Histogram Analysis Method (WHAM), allowing the calculate of the free energy difference between the start and end point (Kumar et al., 1992).

FIGURE 7 | Free-energy profile of an 8-base-pair duplex (3′-ACTGACGT-5′ and 3′-ACGTCAGT-5′) at 312K in a simulation volume of side length 15 units.

Generally, however, we have found that this sophisticated approach is not necessary for oxDNA, and a particularly straightforward umbrella sampling method is built into the standalone oxDNA code. When using VMMC, it is possible to specify discrete order parameters $\lambda(x)$ based on the number of base pairs between user-defined groups of nucleotides. For duplex formation, it is fairly straightforward to iteratively identify a biasing potential that facilitates both the sampling of all states and the rapid transition between fully bound and completely detached configurations.

This biasing potential doesn't need to be fine tuned so that all values of $\lambda(x)$ are equally probable in the biased sample—just good enough to facilitate multiple transitions backwards and forwards. Typical examples for 5-base and 8-base duplexes are given in (Sengar, 2021). For more complex systems, more sophisticated $\lambda(x)$ and the use of multiple sampling windows are sometimes necessary—we refer the reader to (Ouldridge et al., 2010a; Ouldridge et al., 2013a; Machinek et al., 2014; Harrison et al., 2019). Even in these cases, however, the principles are similar to those outlined here.

We perform umbrella sampling simulations on an 8-nucleotide duplex at 312K in a simulation volume of side length 15 units, using the oxDNA1.5 version of the model. Five independent simulations are performed for 7.7×10^8 VMMC steps per particle. The quantity $p_{eq}(Q)$ obtained from simulations is used to calculate a free energy $F(Q)$ according to **Eq. 5** and plotted in **Figure 7**. The shape of this graph is typical for duplex formation, showing the expected large jump in free energy from 0 to 1 base pairs. From 1 to 6 base pairs there is a steady drop in the free energy as configurations are stabilised by additional base-pairing interactions that are favoured once the strands are in close proximity. The final base pairs are less favourable, as base pairs at the end of a duplex are prone to fraying (SantaLucia and Hicks, 2004; Ouldridge et al., 2011).

The shape of $F(Q)$ gives a good guide to constructing first estimates of umbrella biases $W(\lambda(x))$ for duplex formation in general. Ignoring the dissociated state, $W(\lambda(x))$ should increase

roughly exponentially with the number of base pairs broken, since it must counteract $\exp(-F(Q)/kT)$. The slope of $F(Q)$, and hence the required rate of exponential growth in $W(\lambda(x))$, is determined by the temperature; as a crude rule of thumb, a bias of a factor of 10–15 is required per base pair broken at 300K; this required bias falls to a factor of 3-4 by 330 K.

The initial jump in free energy from 0 to 1 in 7 is largely determined by the simulation volume; for simulation cells similar in size to this one, a factor of 3000–10000 is a reasonable first guess for the required weight of the 1-base-pair state relative to the 0-base-pair state.

In addition to biasing by the number of base pairs formed, it is sometimes helpful to also use a distance-based contribution to the order parameter. Built in to the standalone oxDNA code is the ability to define additional dimensions of $\lambda(x)$ that depend on the minimum separation between sets of nucleotides, rather than the number of base pairs. We have found that a simple division of the 0-base-pair state into configurations in which the strands are close (less than 4 units apart) and far apart (4 or more units apart) can reduce the amount of time spent sampling the independent diffusion of strands around the simulation volume. For simulation volumes similar to this one, the close state should be weighted by around 5–10 relative to the distant state, which dominates the unbound ensemble.
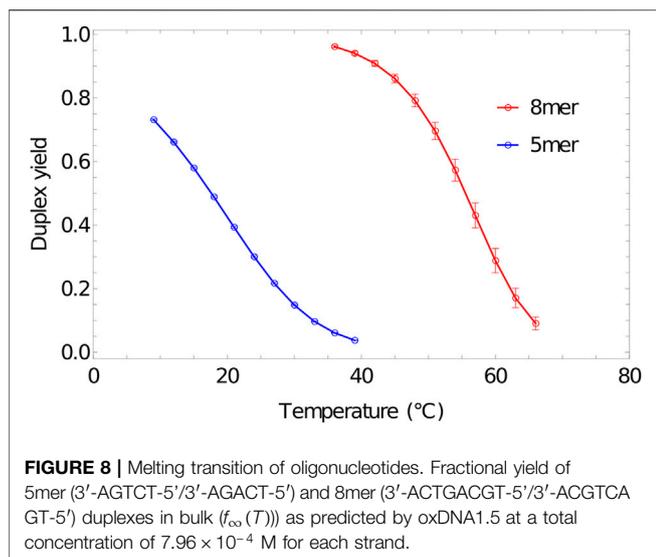
6.1.2 Melting Temperature Curves

Although free-energy profiles for a single pair of strands are informative, they aren't directly comparable to the majority of experiments. Indeed, the thermodynamics of the oxDNA model was parameterised to reproduce the nearest neighbour model (SantaLucia, 1998), which in turn was fitted to—and predicts—experimental melting curves in bulk conditions. The Santalucia parameterisation of the nearest neighbour model (SantaLucia and Hicks, 2004) assumes that DNA duplex formation is essentially a two-state transition between a well-formed duplex and separated single strands. The free-energy profiles produced by oxDNA, such as **Figure 7** are consistent with this picture; the ensemble is dominated by configuration with either zero base pairs, or a large number. In this limit, the melting behaviour can be well-characterised by the fraction of strands that are expected to have base pairing with another strand at a temperature T in a bulk system, $f_\infty(T)$.

To calculate $f_\infty(T)$ in this two-state description, it is first necessary to obtain data at a range of temperatures. In principle, these data can be obtained through separate simulations. However, we have found that a technique called single histogram reweighting (Ferrenberg and Swendsen, 1988) is sufficient to infer $f_\infty(T)$ accurately over a large enough range of temperatures to describe the melting transition. The basic idea is to treat a simulation at a temperature T as a biased sample of the ensemble at another temperature T'; this bias can be corrected in the same way as the bias applied during umbrella sampling:

$$\langle A(x)\rangle_{T'} = \langle A(x)\exp(V_0(x,T)/kT - V_0(x,T')/kT')\rangle_T. \quad (7)$$

Here $V_0(x,T)$ is the value of the potential in the original simulation at temperature T ($V_0(x,T')$ is slightly different due to

FIGURE 8 | Melting transition of oligonucleotides. Fractional yield of 5mer (3′-AGTCT-5′/3′-AGACT-5′) and 8mer (3′-ACTGACGT-5′/3′-ACGTCA GT-5′) duplexes in bulk ($f_\infty(T)$)) as predicted by oxDNA1.5 at a total concentration of 7.96×10^{-4} M for each strand.

a T-dependent term in the potential (Ouldridge et al., 2011)). Extrapolation to nearby temperatures using single histogram reweighting is built into the oxDNA standalone code. It is important to note that if umbrella sampling, and particularly temperature reweighting, are applied, then it is especially important to simulate for a good equilibration time before results are collected. Normally, any initial unrepresentative states will be swiftly overwhelmed within an average taken over the whole course of the simulation. The unbiasing factors in **Eq. 6** and **Eq. 7**, however, can cause unrepresentative initial states to be assigned enormous weights in the ensemble average that are effectively insurmountable, rendering the simulation results meaningless.

Given well-sampled data of the formation of a single duplex in a simulation volume, it is tempting to assume that the fractional yield of states with more than one base pair in a single duplex simulation, $f_1(T)$, is equal to the bulk yield of duplexes $f_\infty(T)$ in a system with the same total concentration of strands. Unfortunately this is not the case; simulations of only a single target duplex neglect concentration fluctuations within unit cells that have large effects on the yield of products (Ouldridge et al., 2010b; Ouldridge, 2012). Quantitative comparison to experimental data is therefore impossible unless extrapolations to bulk conditions can be performed. Assuming ideal behaviour of solutes, Extrapolation is possible. For dimerisation between non-self-complementary strands (Ouldridge et al., 2010b)

$$f_\infty(T) = \left(1 + \frac{1}{2\Phi(T)}\right) - \sqrt{\left(1 + \frac{1}{2\Phi(T)}\right)^2 - 1}, \quad (8)$$

where $\Phi(T) = f_1(T)/(1 - f_1(T))$. A similar result holds for self-complementary duplexes (Ouldridge et al., 2010b), and algorithms exist to extrapolate to bulk for more complex assemblies (Ouldridge, 2012).

Melting curves obtained for 5-base and 8-base duplexes, using umbrella sampling, temperature reweighting and extrapolation to bulk, are reported in **Figure 8**. The melting temperatures T_m for these duplexes – defined, in the two state model, as the temperature at which $f_\infty(T)$ is 0.5 – are close to the values predicted by the nearest neighbour model at the same conditions (17.8°C and 56.1°C) (SantaLucia and Hicks, 2004). This agreement is, of course, due to the model being fitted to these data. However, it is worth noting that although duplexes are often described as having a single "melting temperature", the temperature at which $f_\infty(T)$ is 0.5 depends on the concentration of the individual strands, $[C]$ with (Ouldridge et al., 2011)

$$\frac{dT_m}{d[C]} \sim \frac{\Delta T}{[C]}. \quad (9)$$

Here, ΔT is the width of the transition over which $f_\infty(T)$ goes from largely bound to largely unbound. To match nearest neighbour predictions for melting temperatures over a range of concentrations, therefore, it is necessary that transition widths are also comparable; achieving a good match was a major part of oxDNA's parameterisation.

6.2 Thermodynamics of More Complex Structures

Although accurately simulating basic duplex formation was necessary for the parameterisation of oxDNA, little new information is to be gained from performing these simulations again. The model is trained to reproduce the thermodynamics of the nearest neighbour model, so simulating the thermodynamics of duplex formation is an expensive way to get at an approximation to said nearest neighbour model.Where oxDNA can add value is if duplex formation occurs as part of some more complex system - possibly one in which internally or externally-applied stresses, or topological constraints, are relevant (Romano et al., 2012; Šulc et al., 2012; Ouldridge et al., 2013a; Mosayebi et al., 2015; Kočar et al., 2016; Fonseca et al., 2018; Tee and Wang, 2018; Harrison et al., 2019). As an example, we simulate the formation of a small pseudoknotted structure (**Figure 9**) leveraging the intuition and techniques discussed in **Section 6**. Here, the two sequences 3′-AGCTTCCATG-5′ and 3′-AAGCTCATGG-5′ cannot form a single continuous duplex, but can form two 5-bp duplexes section if both strands bend back on themselves. The stability of this structure cannot be inferred from the nearest neighbour model, but it can easily be simulated with oxDNA. Applying umbrella sampling, we simulate the system at a temperature of 308K in a periodic cell of side-length 20 using oxDNA1.5.

The resulting free-energy profile, **Figure 10**, shows that, at the temperatures of interest, forming two arms is less favourable than forming only one. The advantage obtained by bringing the strand into close proximity via the binding of the first duplex is not enough to overcome the internal stress generated by the structure. This internal stress is evidenced by the much shallower slope of the free energy profile for forming base pairs 6–10 than 1-5.
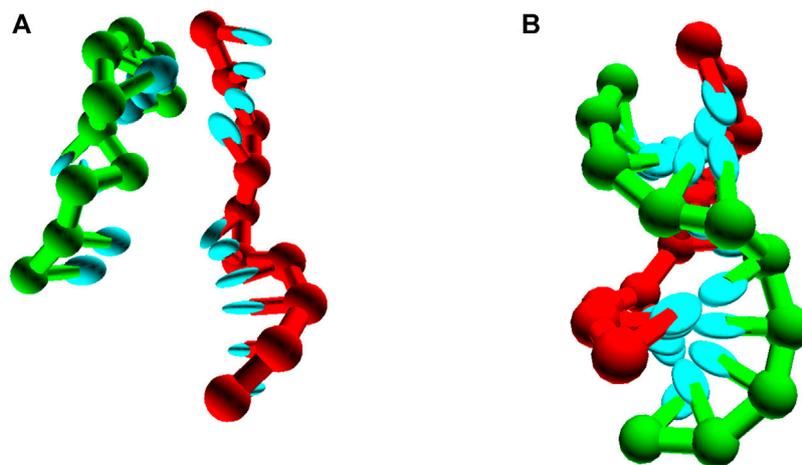
FIGURE 9 | Pseudoknot unbound (A) and bound (B) states, for seqeunces 3′-AGCTTCCATG-5′/3′-AAGCTCATGG-5′.
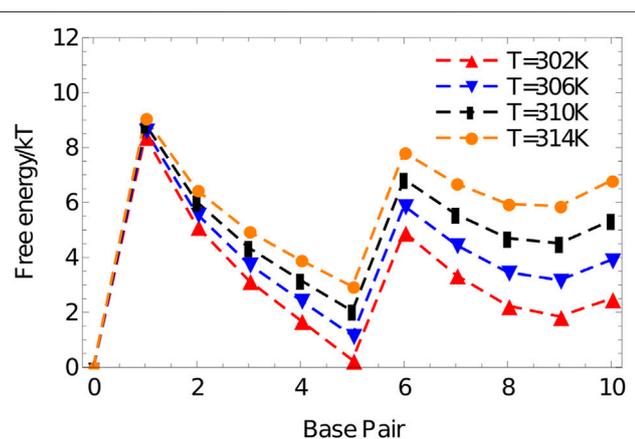
FIGURE 10 | Free energy vs number of base pairs formed for the complex 3′-AGCTTCCATG-5′/3′-AAGCTCATGG-5′ in a simulation volume of side length 20 oxDNA units.

7 DYNAMICAL SIMULATIONS

The simulations described hitherto probe static quantities obtained in the equilibrium ensemble. However, the dynamics of DNA-based systems can be equally important. In particular, the time required for reactions to happen is crucial when constructing complex self-assembling systems or functional circuits, particularly those that are intended to remain out of equilibrium, or exhibit an extremely slow relaxation to equilibrium (Dunn et al., 2015; Srinivas et al., 2017; Fern and Schulman, 2018; Cabello-Garcia et al., 2021).
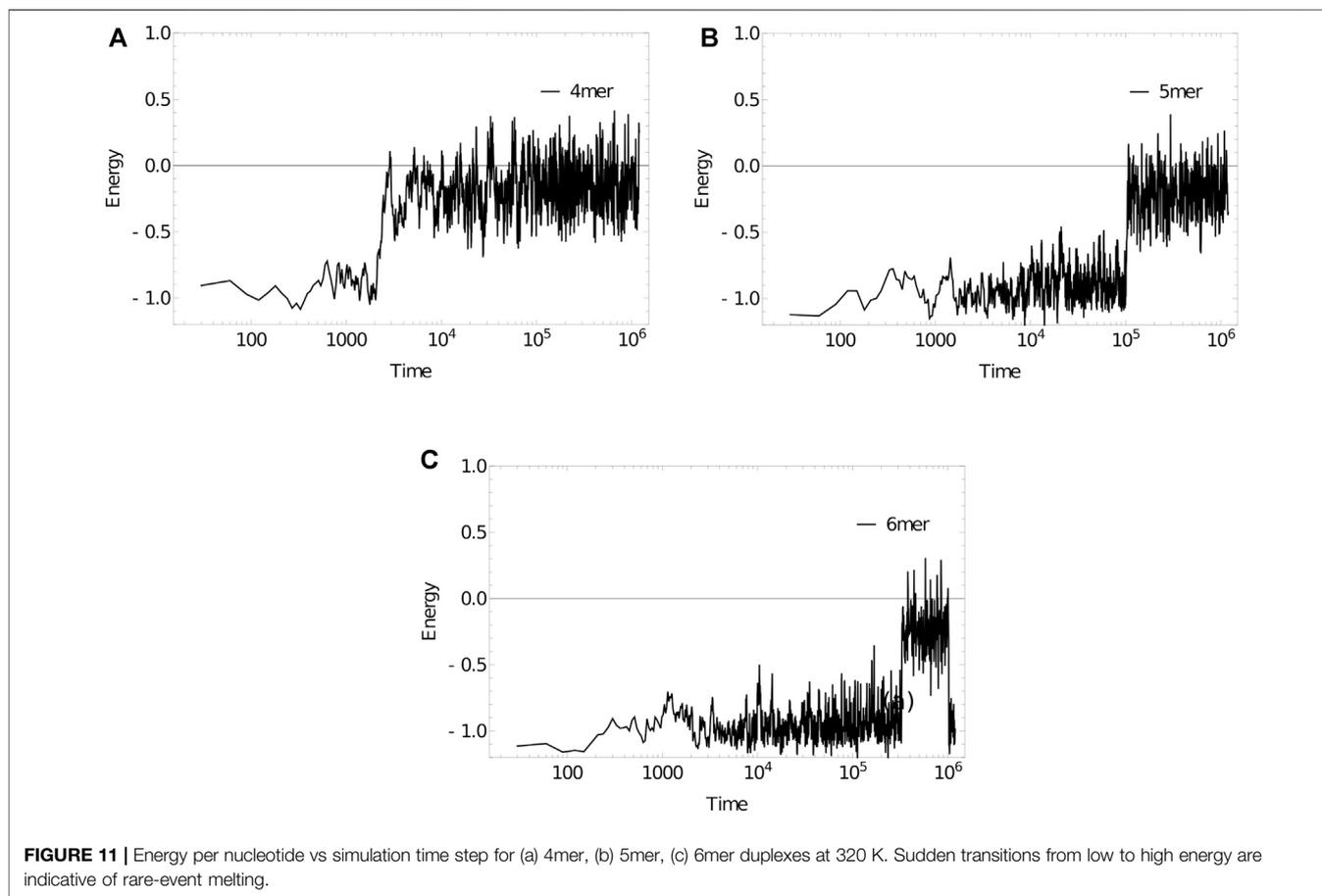
Unlike the thermodynamic and structural properties, oxDNA has not been carefully parameterised to the dynamics of physical DNA. Coarse-graining is generally known to speed up timescales by smoothing free-energy landscapes (Murtola et al., 2009). Moreover, the explicitly dynamical algorithms (particularly the Andersen-like thermostat) give a fairly crude approximation to

the dynamics expected from small molecules in solution. Neither the Anderson-like nor the Langevin thermostat incorporates cooperative hydrodynamics (an updated version of the Langevin thermostat developed to describe hydrodynamic effects (Davidchack et al., 2017) has not yet been implemented in LAMMPS), and both are typically run with large effective diffusion coefficients to enhance sampling (see **Supplementary Appendix A**).

Nonetheless, the dynamics of oxDNA is fundamentally constrained by the combination of its free-energy landscape and its embedding of that free-energy landscape in an explicit geometrical description. For comparison, it is surprisingly difficult to generate meaningful dynamics based on just the free-energy landscape predicted by the nearest neighbour model without an explicit geometrical representation (Srinivas et al., 2013; Schaeffer et al., 2015) [add commented-out citation].

As a result, dynamical simulations of oxDNA can provide useful insight into dynamical properties of physical DNA; the model has been particularly successful in describing toehold-mediated strand displacement (Srinivas et al., 2013; Machinek et al., 2014; Haley et al., 2020; Irmisch et al., 2020), one of the fundamental reactions of DNA nanotechnology. Importantly, the focus should always be on comparing the relative dynamics of two similar systems – for example, the dependence of strand displacement rates on toehold lengths. Unlike the thermodynamic and mechanical properties of oxDNA, absolute values of dynamical properties are largely irrelevant.

As an example, we simulate the dissociation kinetics of duplexes of length 4 (3′-ATAT-5′/3′-ATAT-5′), 5 (3′-ATATA-5′/3′-ATATA-5′) and 6 (3′-ATATAT-5′/3′-ATATAT-5′) at 320 K using the Anderson-like thermostat applied to oxDNA1.5, **Figure 11**. Example trajectories, showing the energy of the system per nucleotide, illustrate the two-state nature of the system discussed earlier in **Section 6**. The strands spend a substantial amount of time in states with an energy of approximately -1.0 in oxDNA units (duplex configurations), before suddenly transitioning to states with an

FIGURE 11 | Energy per nucleotide vs simulation time step for (a) 4mer, (b) 5mer, (c) 6mer duplexes at 320 K. Sudden transitions from low to high energy are indicative of rare-event melting.

energy around -0.2 (single-stranded states). As hinted at by these examples, longer strands take exponentially longer to dissociate (the simulation steps taken to reach an energy of -0.2 per nucleotide, averaged over 10 simulations for each length, are: 9117 ± 2883, 64032 ± 20249, 279744 ± 88463). This exponential suppression of the dissociation rate with strand length is consistent with dissociation being a rare event that requires the crossing of a free energy barrier whose height grows linearly with duplex length at fixed temperature (here, 320 K), as suggested by the free-energy profile in **Figure 7**.

In this case, all systems studied showed the required behaviour on relatively short time scales. Frequently, it is necessary to simulate much slower processes. We have found that the forward flux sampling (FFS) technique (Allen et al., 2009) is an effective tool for simulating dynamical processes with a longer timescale. However, FFS is trickier to implement than umbrella sampling, and is not yet built in to the released code in an optimal way.
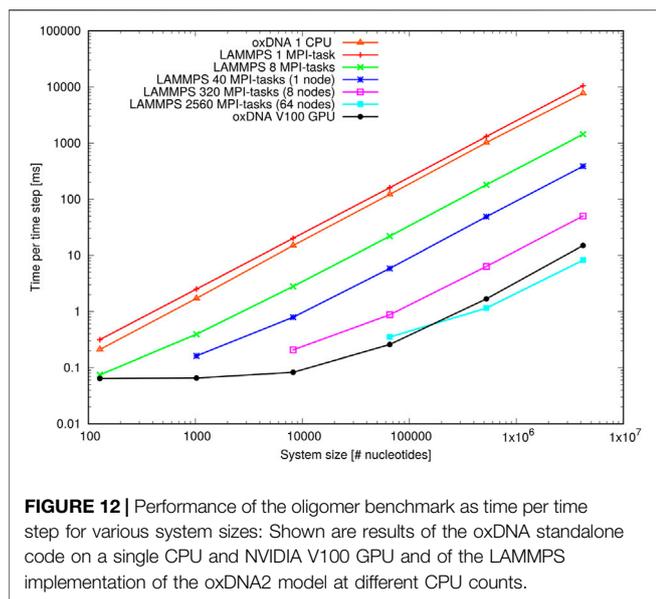
8 Simulation of Large Structures

Another significant application area for oxDNA has been the simulation of large structures to assess their conformation, stability and flexibility (Fernandez-Castanon et al., 2016; Schreck et al., 2016; Sharma et al., 2017; Shi et al., 2017; Benson et al., 2018; Choi et al., 2018; Coronel et al., 2018;

Berengut et al., 2019; Brady et al., 2019; Hoffecker et al., 2019; Snodin et al., 2019; Berengut et al., 2020; Chhabra et al., 2020; Poppleton et al., 2020; Tortora et al., 2020; Yao et al., 2020).

In this context, oxDNA represents an alternative to the CanDo model and simulation package (Castro et al., 2011). The added complexity of oxDNA has a computational cost, but means that it is better able to handle irregular systems. For such simulations, use of oxDNA2.0 is strongly recommended given its better representation of structure, particularly in the context of DNA origami. A more detailed primer on setting up these simulations can be fund in Ref (Doye et al., 2020); here we focus only on technical aspects of the simulations.

As briefly mentioned in **Section 4**, MD algorithms can facilitate the simulation of really large systems by allowing parallelisation across GPU threads or multiple CPUs. The oxDNA standalone code is GPU-enabled via the CUDA C API and supports runs on single CPUs and single GPUs, whereas the LAMMPS version of oxDNA uses the Message Passing Interface (MPI) and is optimised for parallel runs on multi-core CPUs and distributed memory architectures.

To provide benchmarks and examplar codes, we have performed large-scale simulations with both implementations on two different compute architectures, namely a NVIDIA V100 PCIe GPU with 5,120 CUDA cores at Arizona State University's High Performance Computing Facility, and the

FIGURE 12 | Performance of the oligomer benchmark as time per time step for various system sizes: Shown are results of the oxDNA standalone code on a single CPU and NVIDIA V100 GPU and of the LAMMPS implementation of the oxDNA2 model at different CPU counts.

ARCHIE-WeSt HPC facility at the University of Strathclyde consisting of 64 Intel Xeon Gold 6138 (Skylake) processors @2.0GHz with 40 cores per node and 2,560 cores in total. The GPU and single-core CPU runs were performed with the oxDNA standalone code SVN version 6989. The GPU runs all used mixed precision and an edge-based approach (Rovigatti et al., 2015). The LAMMPS stable version from March 3, 2020 was used for the multi-core CPU runs (the small error in the LAMMPS implementation highlighted in **Supplementary Appendix B** is irrelevant for the purposes of these efficiency comparisons). All runs were performed with the oxDNA2.0 model featuring sequence-dependent stacking and hydrogen-bonding interactions.

Two different benchmarks were studied to analyse the performance of both implementations. The first one consisted of a varying number of double-stranded octamer duplexes and investigated the performance at different system sizes, ranging from 8 octamers with 128 nucleotides in total to 262,144 octamers with 4,194,304 nucleotides in total. The concentration of octamers was kept constant at one octamer per 20^3 oxDNA length units, whereas the temperature and salt concentration were set to $T = 293K$ and $[Na^+] = 500$ mM, respectively.

The second benchmark consisted of a DNA origami "pointer" structure (Bai et al., 2012) (15,238 nucleotides) and tested the performance at different salt concentrations between $[Na^+] = 100$ mM and 1 M. The salt concentration is another performance-critical aspect in the simulation of nucleic acids that is often neglected. The reason is that the salt concentration affects the Debye screening length, which is proportional to the inverse square root of the salt concentration. The temperature of this second benchmark was fixed at $T = 293K$. The initial configuration was converted from the cadnano format using the TacoxDNA server (Suma et al., 2019), then relaxed using oxdna.org (Poppleton et al., 2021), implementing the protocol
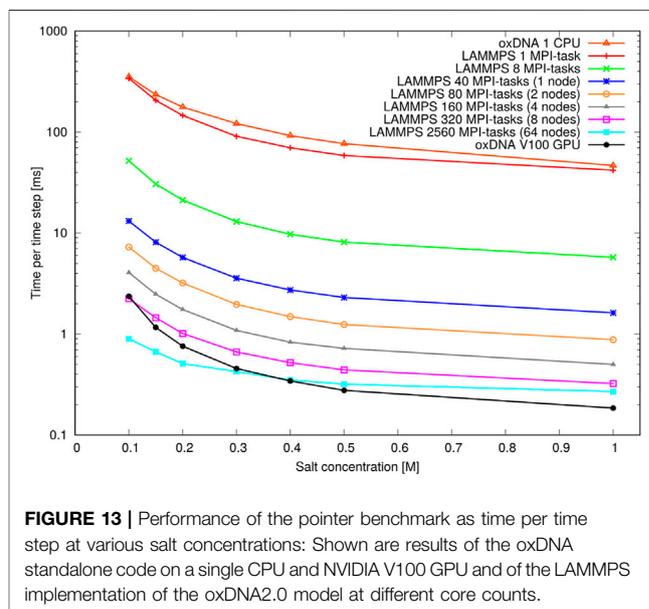
from (Doye et al., 2020) followed by a simulation at the respective salt concentration.

It is worth emphasising that origami structures such as the pointer are a setting in which the improved structural model of oxDNA2.0 is essential. Unless an accurate model is used, relatively small discrepancies can contribute strain that builds up across the structure, resulting in large scale distortion.

Figure 12 shows the results of the oligomer benchmark, which are expressed as time per integration time step in milliseconds. On a single Intel Xeon Gold CPU the standalone code implements a single timestep slightly faster than the LAMMPS implementation. Note, however, that the actual efficiency will depend on the choice of coefficients of coupling to the thermostats (see **Supplementray Appendix B**).

When deployed in parallel on more CPUs, the LAMMPS implementation offsets this disadvantage almost immediately. Its performance at the larger side of system sizes is more or less ideal as evidenced through the linear increase of time per integration step with system size. For smaller system sizes, and depending on how many CPUs were used, the performance levels off due to a build-up of MPI communication overheads. However, there is still a noticeable speed-up e.g. for 8,192 nucleotides on 320 MPI-tasks or 65,536 nucleotides on 2,560 MPI-tasks, which comes down to a very low 25 nucleotides per MPI-task. This unusually good performance of a parallel molecular dynamics code has been reported before (Henrich et al., 2018) and is owed to the rather complex oxDNA force field as the code spends a good deal of time carrying out the force calculation.

The GPU-implementation of the standalone code retains a significant advantage over the LAMMPS implementation for all but the largest benchmark sizes and runs on the full ARCHIE-WeSt system size (2,560 MPI-tasks) and its performance levels only off when the GPU becomes under-subscribed with threads at smaller system sizes. We can conclude that the LAMMPS implementation of oxDNA, besides its capability to run on a

FIGURE 13 | Performance of the pointer benchmark as time per time step at various salt concentrations: Shown are results of the oxDNA standalone code on a single CPU and NVIDIA V100 GPU and of the LAMMPS implementation of the oxDNA2.0 model at different core counts.

variety of CPU architectures, is very suitable for studying small and intermediate system sizes, whereas the GPU-implementation has clearly the edge at large-scale simulations.

Figure 13 shows the performance with the pointer benchmark, again expressed as time per time step in milliseconds. This time the LAMMPS implementation is marginally faster than the oxDNA standalone code on a single CPU. Again, the GPU-runs of the standalone code features significantly shorter run times on all but the largest core counts and lowest salt concentrations. It appears the increase in run time between high and low salt concentration is slightly larger for the GPU-implementation of the standalone code. This could be due to a slightly better handling of neighbour lists in LAMMPS.

Most importantly, however, an increase in runtime by a factor 8–9 can be seen at all core counts when moving from high to moderate salt concentrations. This slowdown is in line with the increase in Debye length by about a factor 3 and reflects the longer cutoff radii and neighbour lists of the pair interactions. This large performance difference should be taken into account when choosing simulation parameters: For instance it is nearly always more convenient to perform relaxation runs to create well-initialized configurations at high salt concentrations (*e.g.* $[Na^+]$ = 1 M). Indeed, unless the response of the system to decreased salt concentration is of specific interest, we would generally recommend using high monovalent salt concentration such as $[Na^+]$ = 1 M for the actual data collection.

9 CONCLUSION

We have reviewed the properties of, and simulation methods available for, the oxDNA model. In the process we have created a well-documented library of examplar simulations available from (Sengar, 2021). Equally importantly, however, we have attempted to provide the necessary intuition both for successfully running oxDNA-based simulations, and also for identifying which systems would actually benefit from those simulations in the first place.

Having explored the model's strengths in some detail, it is worth noting a few natural directions for improvements. Although the model has well-parameterised sequence-dependent thermodynamics,

and a good representation of average mechanical properties, it lacks sequence-dependent structure and mechanics. Incorporating this feature would be useful in and of itself, but would also be a useful first step towards building a model that could interface with other molecules such as proteins (Procyk et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.693710/full#supplementary-material

REFERENCES

Adleman, L. (1994). Molecular Computation of Solutions to Combinatorial Problems. *Science* 266, 1021–1024. doi:10.1126/science.7973651

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *Molecular Biology of the Cell*. 4th ed. (New York: Garland Science).

Allen, R. J., Valeriani, C., and ten Wolde, P. R. (2009). Forward Flux Sampling for Rare Event Simulations. *J. Phys. Condens. Matter* 21, 463102. doi:10.1088/0953-8984/21/46/463102

Bae, J. H., Fang, J. Z., and Zhang, D. Y. (2020). High-throughput Methods for Measuring DNA Thermodynamics. *Nucleic Acids Res.* 48, e89. doi:10.1093/nar/gkaa521

Bai, X.-c., Martin, T. G., Scheres, S. H., and Dietz, H. (2012). Cryo-em Structure of a 3D DNA-Origami Object. *Proc. Natl. Acad. Sci.* 109, 20012–20017. doi:10.1073/pnas.1215713109

Benson, E., Mohammed, A., Rayneau-Kirkhope, D., Gådin, A., Orponen, P., and Högberg, B. (2018). Effects of Design Choices on the Stiffness of Wireframe DNA Origami Structures. *ACS nano* 12, 9291–9299. doi:10.1021/acsnano.8b04148

Berengut, J. F., Berengut, J. C., Doye, J. P., Prešern, D., Kawamoto, A., Ruan, J., et al. (2019). Design and Synthesis of Pleated DNA Origami Nanotubes with Adjustable Diameters. *Nucleic Acids Res.* 47, 11963–11975. doi:10.1093/nar/gkz1056

Berengut, J. F., Wong, C. K., Berengut, J. C., Doye, J. P., Ouldridge, T. E., and Lee, L. K. (2020). Self-limiting Polymerization of DNA Origami Subunits with Strain Accumulation. *ACS nano* 14, 17428–17441. doi:10.1021/acsnano.0c07696

Brady, R. A., Kaufhold, W. T., Brooks, N. J., Foderà, V., and Di Michele, L. (2019). Flexibility Defines Structure in Crystals of Amphiphilic DNA Nanostars. *J. Phys. Condensed Matter* 31, 074003. doi:10.1088/1361-648x/aaf4a1

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). CHARMM: A Program for Macromolecular Energy,

Minimization, and Dynamics Calculations. *J. Comput. Chem.* 4, 187–217. doi:10.1002/jcc.540040211

Brown, R. F., Andrews, C. T., and Elcock, A. H. (2015). Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised Amber Parameters and Compared with experiment. *J. Chem. Theor. Comput.* 11, 2315–2328. doi:10.1021/ct501170h

Brutzer, H., Luzzietti, N., Klaue, D., and Seidel, R. (2010). Energetics at the DNA Supercoiling Transition. *Biophysical J.* 98, 1267–1276. doi:10.1016/j.bpj.2009.12.4292

Bryant, Z., Stone, M. D., Gore, J., Smith, S. B., Cozzarelli, N. R., and Bustamante, C. (2003). Structural Transitions and Elasticity from Torque Measurements on DNA. *Nature* 424, 338–341. doi:10.1038/nature01810

Cabello-Garcia, J., Bae, W., Stan, G.-B. V., and Ouldridge, T. E. (2021). Handhold-mediated Strand Displacement: A Nucleic Acid Based Mechanism for Generating Far-From-Equilibrium Assemblies through Templated Reactions. *ACS nano* 15, 3272–3283. doi:10.1021/acsnano.0c10068

Cantor, C. R., and Schimmel, P. R. (1980). *Biophysical Chemistry: Part III: The Behavior of Biological Macromolecules*. San Francisco: Macmillan.

Caraglio, M., Skoruppa, E., and Carlon, E. (2019). Overtwisting Induces Polygonal Shapes in Bent DNA. *J. Chem. Phys.* 150, 135101. doi:10.1063/1.5084950

Castro, C. E., Kilchherr, F., Kim, D. N., Shiao, E. L., Wauer, T., Wortmann, P., et al. (2011). A Primer to Scaffolded DNA Origami. *Nat. Meth.* 8, 221–229. doi:10.1038/nmeth.1570

Chakraborty, D., Hori, N., and Thirumalai, D. (2018). Sequence-dependent Three Interaction Site Model for Single-And Double-Stranded DNA. *J. Chem. Theor. Comput.* 14, 3763–3779. doi:10.1021/acs.jctc.8b00091

Chen, H., Meisburger, S. P., Pabit, S. A., Sutton, J. L., Webb, W. W., and Pollack, L. (2012). Ionic Strength-dependent Persistence Lengths of Single-Stranded RNA and DNA. *Proc. Natl. Acad. Sci. United States America* 109, 799–804. doi:10.1073/pnas.1119057109

Cherry, K. M., and Qian, L. (2018). Scaling up Molecular Pattern Recognition with DNA-Based winner-take-all Neural Networks. *Nature* 559, 370–376. doi:10.1038/s41586-018-0289-6

Chhabra, H., Mishra, G., Cao, Y., Presern, D., Skoruppa, E., Tortora, M. M., et al. (2020). Computing the Elastic Mechanical Properties of Rodlike DNA Nanostructures. *J. Chem. Theor. Comput.* 16, 7748–7763. doi:10.1021/acs.jctc.0c00661

Choi, Y., Choi, H., Lee, A. C., and Kwon, S. (2018). Design and Synthesis of a Reconfigurable DNA Accordion Rack. *J. Vis. Exp.* 15, 58364. doi:10.3791/58364

Cocco, S., and Monasson, R. (1999). Statistical Mechanics of Torque Induced Denaturation of DNA. *Phys. Rev. Lett.* 83, 5178–5181. doi:10.1103/PhysRevLett.83.5178

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., et al. (1996). A Second Generation Force Field for the Simulation of Proteins. *Nucleic Acids Org. Mol. J. Am. Chem. Soc.* 117118, 5179–5197. doi:10.1021/ja955032e

Coronel, L., Suma, A., and Micheletti, C. (2018). Dynamics of Supercoiled DNA with Complex Knots: Large-Scale Rearrangements and Persistent Multi-Strand Interlocking. *Nucleic Acids Res.* 46, 7533–7541. doi:10.1093/nar/gky523

Craggs, T. D., Sustarsic, M., Plochowietz, A., Mosayebi, M., Kaju, H., Cuthbert, A., et al. (2019). Substrate Conformational Dynamics Facilitate Structure-specific Recognition of Gapped DNA by DNA Polymerase. *Nucleic Acids Res.* 47, 10788–10800. doi:10.1093/nar/gkz797

Crothers, D. M., Drak, J., Kahn, J. D., and Levene, S. D. (1992). DNA Bending, Flexibility, and Helical Repeat by Cyclization Kinetics. *Methods Enzymol.* 212, 3–29. doi:10.1016/0076-6879(92)12003-9

Dans, P. D., Walther, J., Gómez, H., and Orozco, M. (2016). Multiscale Simulation of DNA. *Curr. Opin. Struct. Biol.* 37, 29–45. doi:10.1016/j.sbi.2015.11.011

Dauxois, T., Peyrard, M., and Bishop, A. R. (1993). Dynamics and Thermodynamics of a Nonlinear Model for DNA Denaturation. *Phys. Rev. E* 47, 684–695. doi:10.1103/PhysRevE.47.684

Davidchack, R. L., Ouldridge, T. E., and Tretyakov, M. V. (2017). Geometric Integrator for Langevin Systems with Quaternion-Based Rotational Degrees of freedom and Hydrodynamic Interactions. *J. Chem. Phys.* 147, 224103. doi:10.1063/1.4999771

Davidchack, R. L., Ouldridge, T. E., and Tretyakov, M. V. (2015). New Langevin and Gradient Thermostats for Rigid Body Dynamics. *J. Chem. Phys.* 142 (14), 144114.

De Michele, C., Rovigatti, L., Bellini, T., and Sciortino, F. (2012). Self-assembly of Short DNA Duplexes: From a Coarse-Grained Model to Experiments through a Theoretical Link. *Soft Matter* 8, 8388–8398. doi:10.1039/c2sm25845e

Demurtas, D., Amzallag, A., Rawdon, E. J., Maddocks, J. H., Dubochet, J., and Stasiak, A. (2009). Bending Modes of DNA Directly Addressed by Cryo-Electron Microscopy of DNA Minicircles. *Nucleic Acids Res.* 37, 2882–2893. doi:10.1093/nar/gkp137

Desai, P. R., Brahmachari, S., Marko, J. F., Das, S., and Neuman, K. C. (2020). Coarse-grained Modelling of DNA Plectoneme Pinning in the Presence of Base-Pair Mismatches. *Nucleic Acids Res.* 48, 10713–10725. doi:10.1093/nar/gkaa836

Dessinges, M.-N., Maier, B., Zhang, Y., Peliti, M., Bensimon, D., and Croquette, V. (2002). Stretching Single Stranded DNA, a Model Polyelectrolyte. *Phys. Rev. Lett.* 89, 248102. doi:10.1103/physrevlett.89.248102

Douglas, S. M., Dietz, H., Liedl, T., Högberg, B., Graf, F., and Shih, W. M. (2009). Self-assembly of DNA into Nanoscale Three-Dimensional Shapes. *Nature* 459, 414–418. doi:10.1038/nature08016

Doye, J. P. K., Fowler, H., Prešern, D., Bohlin, J., Rovigatti, L., Romano, F., et al. (2020). *The oxDNA Coarse-Grained Model as a Tool to Simulate DNA Origami* (arXiv:2004.05052). Available at: https://www.arxiv.org/abs/2004.05052 (Accessed April 20, 2021).

Doye, J. P. K., Ouldridge, T. E., Louis, A. A., Romano, F., Šulc, P., Matek, C., et al. (2013). Coarse-graining DNA for Simulations of. *DNA nanotechnology* 15, 20395–20414. doi:10.1039/c3cp53545b

Du, Q., Kotlyar, A., and Vologodskii, A. (2008). Kinking the Double helix by Bending Deformation. *Nucleic Acids Res.* 36, 1120–1128. doi:10.1093/nar/gkm1125

Dunn, K. E., Dannenberg, F., Ouldridge, T. E., Kwiatkowska, M., Turberfield, A. J., and Bath, J. (2015). Guiding the Folding Pathway of DNA Origami. *Nature* 525, 82–86. doi:10.1038/nature14860

Engel, M. C., Romano, F., Louis, A. A., and Doye, J. P. (2020). Measuring Internal Forces in Single-Stranded DNA: Application to a DNA Force Clamp. *J. Chem. Theor. Comput.* 16, 7764–7775. doi:10.1021/acs.jctc.0c00286

Engel, M. C., Smith, D. M., Jobst, M. A., Sajfutdinow, M., Liedl, T., Romano, F., et al. (2018). Force-induced Unravelling of DNA Origami. *ACS nano* 12, 6734–6747. doi:10.1021/acsnano.8b01844

Fern, J., and Schulman, R. (2018). Modular DNA Strand-Displacement Controllers for Directing Material Expansion. *Nat. Commun.* 9, 1–8. doi:10.1038/s41467-018-06218-w

Fernandez-Castanon, J., Bomboi, F., Rovigatti, L., Zanatta, M., Paciaroni, A., Comez, L., et al. (2016). Small-angle Neutron Scattering and Molecular Dynamics Structural Study of Gelling DNA Nanostars. *J. Chem. Phys.* 145, 084910. doi:10.1063/1.4961398

Ferrenberg, A. M., and Swendsen, R. H. (1988). New Monte Carlo Technique for Studying Phase Transitions. *Phys. Rev. Lett.* 61, 2635–2638. doi:10.1103/physrevlett.61.2635

Fields, A. P., Meyer, E. A., and Cohen, A. E. (2013). Euler Buckling and Nonlinear Kinking of Double-Stranded DNA. *Nucleic Acids Res.* 41, 9881–9890. doi:10.1093/nar/gkt739

Flamm, C., Fontana, W., Hofacker, I. L., and Schuster, P. (2000). RNA Folding at Elementary Step Resolution. *RNA* 6, 325–338. doi:10.1017/s1355838200992161

Fonseca, P., Romano, F., Schreck, J. S., Ouldridge, T. E., Doye, J. P., and Louis, A. A. (2018). Multi-scale Coarse-Graining for the Study of Assembly Pathways in DNA-brick Self-Assembly. *J. Chem. Phys.* 148, 134910. doi:10.1063/1.5019344

Forth, S., Deufel, C., Sheinin, M. Y., Daniels, B., Sethna, J. P., and Wang, M. D. (2008). Abrupt Buckling Transition Observed during the Plectoneme Formation of Individual DNA Molecules. *Phys. Rev. Lett.* 100, 148301. doi:10.1103/physrevlett.100.148301

Fosado, Y. A. G., Landuzzi, F., and Sakaue, T. (2021). Twist Dynamics and Buckling Instability of Ring DNA: the Effect of Groove Asymmetry and Anisotropic Bending. *Soft Matter* 17, 1530–1537. doi:10.1039/d0sm01812k

Frenkel, D., and Smit, B. (2002). *Understanding Molecular Simulation: From Algorithms to Applications*. 2 edn. (Orlando, FL: Academic Press).

Fu, T. J., and Seeman, N. C. (1993). DNA Double-Crossover Molecules. *Biochemistry* 32, 3211. doi:10.1021/bi00064a003

Fujimoto, B. S., Brewood, G. P., and Schurr, J. M. (2006). Torsional Rigidities of Weakly Strained Dnas. *Biophysical J.* 91, 4166–4179. doi:10.1529/biophysj.106.087593

Fye, R. M., and Benham, C. J. (1999). Exact Method for Numerically Analyzing a Model of Local Denaturation in Superhelically Stressed DNA. *Phys. Rev. E* 59, 3408–3426. doi:10.1103/physreve.59.3408

Goodman, R. P., Sharp, I. A. T., Tardin, C. F., Erben, C. M., Berry, R. M., Schmidt, C. F., et al. (2005). Rapid Chiral Assembly of Rigid DNA Building Blocks for Molecular Nanofabrication. *Science* 310, 1661–1665. doi:10.1126/science.1120367

Gore, J., Bryant, Z., Nöllmann, M., Le, M. U., Cozzarelli, N. R., and Bustamante, C. (2006). DNA Overwinds when Stretched. *Nature* 442, 836–839. doi:10.1038/nature04974

Gross, P., Laurens, N., Oddershede, L. B., Bockelmann, U., Peterman, E. J., and Wuite, G. J. (2011). Quantifying How DNA Stretches, Melts and Changes Twist under Tension. *Nat. Phys.* 7, 731–736. doi:10.1038/nphys2002

Haley, N. E., Ouldridge, T. E., Ruiz, I. M., Geraldini, A., Louis, A. A., Bath, J., et al. (2020). Design of Hidden Thermodynamic Driving for Non-equilibrium Systems via Mismatch Elimination during DNA Strand Displacement. *Nat. Commun.* 11, 1–11. doi:10.1038/s41467-020-16353-y

Harrison, R. M., Romano, F., Ouldridge, T. E., Louis, A. A., and Doye, J. P. (2015). *Coarse-grained Modelling of strong DNA Bending I: Thermodynamics and Comparison to an Experimental" Molecular Vice"* (arXiv preprint arXiv:1506.09005). Available at: https://arxiv.org/abs/1506.09005 (Accessed April 20, 2021).

Harrison, R. M., Romano, F., Ouldridge, T. E., Louis, A. A., and Doye, J. P. K. (2019). Identifying Physical Causes of Apparent Enhanced Cyclization of Short DNA Molecules with a Coarse-Grained Model. *J. Chem. Theor. Comput.* 15, 4660–4672. doi:10.1021/acs.jctc.9b00112

Hatch, K., Danilowicz, C., Coljee, V., and Prentiss, M. (2008). Demonstration that the Shear Force Required to Separate Short Double-Stranded DNA Does Not Increase Significantly with Sequence Length for Sequences Longer Than 25 Base Pairs. *Phys. Rev. E* 78, 011920. doi:10.1103/PhysRevE.78.011920

Henning-Knechtel, A., Knechtel, J., and Magzoub, M. (2017). NAR Breakthrough Article DNA-Assisted Oligomerization of Pore-Forming Toxin Monomers into Precisely-Controlled Protein Channels. *Nucleic Acids Res.* 45, 12057–12068. doi:10.1093/nar/gkx990

Henrich, O., Gutiérrez Fosado, Y. A., Curk, T., and Ouldridge, T. E. (2018). Coarse-grained Simulation of DNA Using LAMMPS: An Implementation of the oxDNA Model and its Applications. *Eur. Phys. J. E* 41, 1–16. doi:10.1140/epje/i2018-11669-8

Herrero-Galán, E., Fuentes-Perez, M. E., Carrasco, C., Valpuesta, J. M., Carrascosa, J. L., Moreno-Herrero, F., et al. (2013). Mechanical Identities of RNA and DNA Double Helices Unveiled at the Single-Molecule Level. *J. Am. Chem. Soc.* 135, 122–131. doi:10.1021/ja3054755

Hinckley, D. M., Freeman, G. S., Whitmer, J. K., and De Pablo, J. J. (2013). An Experimentally-Informed Coarse-Grained 3-Site-Per-Nucleotide Model of Dna: Structure, Thermodynamics, and Dynamics of Hybridization. *J. Chem. Phys.* 139, 144903. doi:10.1063/1.4822042

Hobza, P., and Šponer, J. (1999). Structure, Energetics, and Dynamics of the Nucleic Acid Base Pairs: Nonempirical Ab Initio Calculations. *Chem. Rev.* 99, 3247–3276. doi:10.1021/cr9800255

Hoffecker, I. T., Chen, S., Gådin, A., Bosco, A., Teixeira, A. I., and Högberg, B. (2019). Solution-controlled Conformational Switching of an Anchored Wireframe DNA Nanostructure. *Small* 15, 1803628. doi:10.1002/smll.201803628

Hong, F., Jiang, S., Lan, X., Narayanan, R. P., Šulc, P., Zhang, F., et al. (2018). Layered-Crossover Tiles with Precisely Tunable Angles for 2D and 3D DNA Crystal Engineering. *J. Am. Chem. Soc.* 140, 14670–14676. doi:10.1021/jacs.8b07180

Huguet, J. M., Bizarro, C. V., Forns, N., Smith, S. B., Bustamante, C., and Ritort, F. (2010). Single-molecule Derivation of Salt Dependent Base-Pair Free Energies in DNA. *Proc. Natl. Acad. Sci. U.S.A.* 107. doi:10.1073/pnas.1001454107

Irmisch, P., Ouldridge, T. E., and Seidel, R. (2020). Modeling DNA-Strand Displacement Reactions in the Presence of Base-Pair Mismatches. *J. Am. Chem. Soc.* 142, 11451–11463. doi:10.1021/jacs.0c03105

Ivani, I., Dans, P. D., Noy, A., Pérez, A., Faustino, I., Hospital, A., et al. (2016). Parmbsc1: a Refined Force Field for DNA Simulations. *Nat. Methods* 13, 55. doi:10.1038/nmeth.3658

Ke, Y., Ong, L. L., Shih, W. M., and Yin, P. (2012). Three-dimensional Structures Self-Assembled from DNA Bricks. *Science* 338, 1177–1183. doi:10.1126/science.1227268

Kim, C., Lee, O.-c., Kim, J.-Y., Sung, W., and Lee, N. K. (2015). Dynamic Release of Bending Stress in Short Dsdna by Formation of a Kink and forks. *Angew. Chem.* 127, 9071–9075. doi:10.1002/ange.201502055

Kočar, V., Schreck, J. S., Čeru, S., Gradišar, H., Bašić, N., Pisanski, T., et al. (2016). Design Principles for Rapid Folding of Knotted DNA Nanostructures. *Nat. Commun.* 7, 1–8. doi:10.1038/ncomms10803

Korolev, N., Luo, D., Lyubartsev, A. P., and Nordenskiöld, L. (2014). A Coarse-Grained DNA Model Parameterized from Atomistic Simulations by Inverse Monte Carlo. *Polymers* 6, 1655–1675. doi:10.3390/polym6061655

Kratky, O., and Porod, G. (1949). Röntgenuntersuchung Gelöster Fadenmoleküle. *Recueil des Travaux Chimiques des Pays-Bas* 68, 1106–1122. doi:10.1002/recl.19490681203

Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1992). The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* 13, 1011–1021. doi:10.1002/jcc.540130812

LAMMPS Documentation (2021). USER-CGDNA package. Available at: https://docs.lammps.org/ (Accessed May 27, 2021).

Le, T. T., and Kim, H. D. (2014). Probing the Elastic Limit of DNA Bending. *Nucleic Acids Res.* 42, 10786–10794. doi:10.1093/nar/gku735

Lee, J. Y., Terakawa, T., Qi, Z., Steinfeld, J. B., Redding, S., Kwon, Y. H., et al. (2015). Base Triplet Stepping by the Rad51/RecA Family of Recombinases. *Science* 349, 977–981. doi:10.1126/science.aab2666

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., et al. (1996). Crystal Structure of the Lactose Operon Repressor and its Complexes with DNA and Inducer. *Science* 271, 1247–1254. doi:10.1126/science.271.5253.1247

Lionnet, T., Joubaud, S., Lavery, R., Bensimon, D., and Croquette, V. (2006). Wringing Out DNA. *Phys. Rev. Lett.* 96, 178102. doi:10.1103/physrevlett.96.178102

Machado, M. R., and Pantano, S. (2015). Exploring Laci–DNA Dynamics by Multiscale Simulations Using the SIRAH Force Field. *J. Chem. Theor. Comput.* 11, 5012–5023. doi:10.1021/acs.jctc.5b00575

Machinek, R. R., Ouldridge, T. E., Haley, N. E., Bath, J., and Turberfield, A. J. (2014). Programmable Energy Landscapes for Kinetic Control of DNA Strand Displacement. *Nat. Commun.* 5, 1–9. doi:10.1038/ncomms6324

Maciejczyk, M., Spasic, A., Liwo, A., and Scheraga, H. A. (2014). DNA Duplex Formation with a Coarse-Grained Model. *J. Chem. Theor. Comput.* 10, 5020–5035. doi:10.1021/ct4006689

Maffeo, C., and Aksimentiev, A. (2020). MrDNA: a Multi-Resolution Model for Predicting the Structure and Dynamics of DNA Systems. *Nucleic Acids Res.* 48, 5135–5146. doi:10.1093/nar/gkaa200

Maffeo, C., Ngo, T. T., Ha, T., and Aksimentiev, A. (2014). A Coarse-Grained Model of Unstructured Single-Stranded DNA Derived from Atomistic Simulation and Single-Molecule experiment. *J. Chem. Theor. Comput.* 10, 2891–2896. doi:10.1021/ct500193u

Matek, C., Ouldridge, T. E., Doye, J. P. K., and Louis, A. A. (2015). Plectoneme Tip Bubbles: Coupled Denaturation and Writhing in Supercoiled DNA. *Sci. Rep.* 5, 7655. doi:10.1038/srep07655

Matek, C., Ouldridge, T. E., Levy, A., Doye, J. P. K., and Louis, A. A. (2012). DNA Cruciform Arms Nucleate through a Correlated but Asynchronous Cooperative Mechanism. *J. Phys. Chem. B* 116, 11616–11625. doi:10.1021/jp3080755

Meng, H., Bosman, J., van der Heijden, T., and van Noort, J. (2014). Coexistence of Twisted, Plectonemic, and Melted DNA in Small Topological Domains. *Biophysical J.* 106, 1174–1181. doi:10.1016/j.bpj.2014.01.017

Metropolis, N., and Ulam, S. (1949). The Monte Carlo Method. *J. Am. Stat. Assoc.* 44, 335–341. doi:10.1080/01621459.1949.10483310

Mills, J. B., Vacano, E., and Hagerman, P. J. (1999). Flexibility of Single-Stranded DNA: Use of Gapped Duplex Helices to Determine the Persistence Lengths of Poly(dT) and Poly(dA). *J. Mol. Biol.* 285, 245–257. doi:10.1006/jmbi.1998.2287

Mosayebi, M., Louis, A. A., Doye, J. P., and Ouldridge, T. E. (2015). Force-induced Rupture of a DNA Duplex: from Fundamentals to Force Sensors. *ACS nano* 9, 11993–12003. doi:10.1021/acsnano.5b04726

Mosconi, F., Allemand, J. F., Bensimon, D., and Croquette, V. (2009). Measurement of the Torque on a Single Stretched and Twisted DNA Using

Magnetic Tweezers. *Phys. Rev. Lett.* 102, 078301. doi:10.1103/PhysRevLett.102.078301

Murphy, M. C., Rasnik, I., Chang, W., Lohman, T. M., and Ha, T. (2004). Probing Single-Stranded DNA Conformational Flexibility Using Flourescence Spectroscopy. *Biophys. J.* 86, 2530–2537. doi:10.1016/s0006-3495(04)74308-8

Murtola, T., Bunker, A., Vattulainen, I., Deserno, M., and Karttunen, M. (2009). Multiscale Modeling of Emergent Materials: Biological and Soft Matter. *Phys. Chem. Chem. Phys.* 11, 1869–1892. doi:10.1039/b818051b

Muscat, R. A., Bath, J., and Turberfield, A. J. (2011). A Programmable Molecular Robot. *Nano Lett.* 11, 982–987. doi:10.1021/nl1037165

Nguyen, H., Maier, J., Huang, H., Perrone, V., and Simmerling, C. (2014). Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. *J. Am. Chem. Soc.* 136, 13959–13962. doi:10.1021/ja5032776

Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G., and Burley, S. K. (1996). Crystal Structure of a Human Tata Box-Binding Protein/tata Element Complex. *Proc. Natl. Acad. Sci.* 93, 4862–4867. doi:10.1073/pnas.93.10.4862

Nisoli, C., and Bishop, A. R. (2011). Thermomechanics of DNA: Theory of thermal Stability under Load. *Phys. Rev. Lett.* 107, 068102. doi:10.1103/PhysRevLett.107.068102

Nomidis, S. K., Skoruppa, E., Carlon, E., and Marko, J. F. (2019). Twist-bend Coupling and the Statistical Mechanics of the Twistable Wormlike-Chain Model of DNA: Perturbation Theory and beyond. *Phys. Rev. E* 99, 032414. doi:10.1103/PhysRevE.99.032414

Odijk, T. (1995). Stiff Chains and Filaments under Tension. *Macromolecules* 28, 7016–7018. doi:10.1021/ma00124a044

Ouldridge, T. E., Hoare, R. L., Louis, A. A., Doye, J. P. K., Bath, J., and Turberfield, A. J. (2013a). Optimizing DNA Nanotechnology through Coarse-Grained Modeling: A Two-Footed DNA walker. *ACS Nano* 7, 2479–2490. doi:10.1021/nn3058483

Ouldridge, T. E. (2012). Inferring Bulk Self-Assembly Properties from Simulations of Small Systems with Multiple Constituent Species and Small Systems in the Grand Canonical Ensemble. *J. Chem. Phys.* 137, 144105. doi:10.1063/1.4757267

Ouldridge, T. E., Johnston, I. G., Louis, A. A., and Doye, J. P. K. (2009). The Self-Assembly of DNA Holliday Junctions Studied with a Minimal Model. *J. Chem. Phys.* 130, 065101. doi:10.1063/1.3055595

Ouldridge, T. E., Louis, A. A., and Doye, J. P. K. (2010a). DNA Nanotweezers Studied with a Coarse-Grained Model of DNA. *Phys. Rev. Lett.* 104, 178101. doi:10.1103/PhysRevLett.104.178101

Ouldridge, T. E., Louis, A. A., and Doye, J. P. K. (2010b). Extracting Bulk Properties of Self-Assembling Systems from Small Simulations. *J. Phys. Condensed Matter* 22, 104102. doi:10.1088/0953-8984/22/10/104102

Ouldridge, T. E., Louis, A. A., and Doye, J. P. K. (2011). Structural, Mechanical, and Thermodynamic Properties of a Coarse-Grained DNA Model. *J. Chem. Phys.* 134, 085101. doi:10.1063/1.3552946

Ouldridge, T. E., Šulc, P., Romano, F., Doye, J. P. K., and Louis, A. A. (2013b). DNA Hybridization Kinetics: Zippering, Internal Displacement and Sequence Dependence. *Nucleic Acids Res.* 41, 8886–8895. doi:10.1093/nar/gkt687

oxDNA wiki (2015). Available at: https://dna.physics.ox.ac.uk/ (Accessed April 20, 2021).

Park, G., Cho, M. K., and Jung, Y. (2021). Sequence-dependent Kink Formation in Short DNA Loops: Theory and Molecular Dynamics Simulations. *J. Chem. Theor. Comput.* 17, 1308–1317. doi:10.1021/acs.jctc.0c01116

Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., III, Laughton, C. A., et al. (2007). Refinement of the Amber Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophysical J.* 92, 3817–3829. doi:10.1529/biophysj.106.097782

Pérez, A., Noy, A., Lankas, F., Luque, F. J., and Orozco, M. (2004). The Relative Flexibility of B-DNA and A-RNA Duplexes: Database Analysis. *Nucleic Acids Res.* 32, 6144–6151. doi:10.1093/nar/gkh954

Podtelezhnikov, A. A., Mao, C., Seeman, N. C., and Vologodskii, A. (2000). Multimerization-cyclization of DNA Fragments as a Method of Conformational Analysis. *Biophysical J.* 79, 2692–2704. doi:10.1016/s0006-3495(00)76507-6

Poland, D., and Scheraga, H. A. (1966). Occurrence of a Phase Transition in Nucleic Acid Models. *J. Chem. Phys.* 45, 1464–1469. doi:10.1063/1.1727786

Poppleton, E., Bohlin, J., Matthies, M., Sharma, S., Zhang, F., and Šulc, P. (2020). Design, Optimization and Analysis of Large DNA and RNA Nanostructures through Interactive Visualization, Editing and Molecular Simulation. *Nucleic Acids Res.* 48, E72. doi:10.1093/nar/gkaa417

Poppleton, E., Romero, R., Mallya, A., Rovigatti, L., and Šulc, P. (2021). Oxdna. Org: a Public Webserver for Coarse-Grained Simulations of DNA and RNA Nanostructures. *Nucleic Acids Res.*

Procyk, J., Poppleton, E., and Šulc, P. (2020). Coarse-grained Nucleic Acid-Protein Model for Hybrid Nanotechnology. *Soft Matter* 17, 3586. doi:10.1039/d0sm01639j

Qian, L., Winfree, E., and Bruck, J. (2011). Neural Network Computation with DNA Strand Displacement Cascades. *Nature* 475, 368–372. doi:10.1038/nature10262

Qu, H., and Zocchi, G. (2011). The Complete Bending Energy Function for Nicked DNA. *EPL (Europhysics Letters)* 94, 18003. doi:10.1209/0295-5075/94/18003

Richmond, T. J., and Davey, C. A. (2003). The Structure of DNA in the Nucleosome Core. *Nature* 423, 145–150. doi:10.1038/nature01595

Rivetti, C., Walker, C., and Bustamante, C. (1998). Polymer Chain Statistics and Conformational Analysis of DNA Molecules with Bends or Sections of Different Flexibility. *J. Mol. Biol.* 280, 41–59. doi:10.1006/jmbi.1998.1830

Rocklin, G. J., Chidyausiku, T. M., Goreshnik, I., Ford, A., Houliston, S., Lemak, A., et al. (2017). Global Analysis of Protein Folding Using Massively Parallel Design, Synthesis, and Testing. *Science* 357, 168–175. doi:10.1126/science.aan0693

Romano, F., Chakraborty, D., Doye, J. P. K., Ouldridge, T. E., and Louis, A. A. (2013). Coarse-grained Simulations of DNA Overstretching. *J. Chem. Phys.* 138, 085101. doi:10.1063/1.4792252

Romano, F., Hudson, A., Doye, J. P. K., Ouldridge, T. E., and Louis, A. A. (2012). The Effect of Topology on the Structure and Free Energy Landscape of DNA Kissing Complexes. *J. Chem. Phys.* 136, 215102. doi:10.1063/1.4722203

Rothemund, P. W. (2006). Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* 440, 297–302. doi:10.1038/nature04586

Rothemund, P. W., Papadakis, N., and Winfree, E. (2004). Algorithmic Self-Assembly of DNA Sierpinski Triangles. *Plos Biol.* 2, e424. doi:10.1371/journal.pbio.0020424

Rovigatti, L., Bomboi, F., and Sciortino, F. (2014). Accurate Phase Diagram of Tetravalent DNA Nanostars. *J. Chem. Phys.* 140, 154903. doi:10.1063/1.4870467

Rovigatti, L., Sulc, P., Reguly, I. Z., and Romano, F. (2015). A Comparison between Parallelization Approaches in Molecular Dynamics Simulations on GPUs. *J. Comput. Chem.* 36, 1–8. doi:10.1002/jcc.23763

Russo, J., Tartaglia, P., and Sciortino, F. (2009). Reversible Gels of Patchy Particles: Role of the Valence. *J. Chem. Phys.* 131, 014504. doi:10.1063/1.3153843

Salerno, D., Tempestini, A., Mai, I., Brogioli, D., Ziano, R., Cassina, V., et al. (2012). Single-molecule Study of the DNA Denaturation Phase Transition in the Force-Torsion Space. *Phys. Rev. Lett.* 109, 118303. doi:10.1103/physrevlett.109.118303

SantaLucia, J. (1998). A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics. *Proc. Natl. Acad. Sci. United States America* 95, 1460–1465. doi:10.1073/pnas.95.4.1460

SantaLucia, J., and Hicks, D. (2004). The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* 33, 415–440. doi:10.1146/annurev.biophys.32.110601.141800

Savelyev, A. (2012). Do monovalent mobile Ions Affect Dna's Flexibility at High Salt Content?. *Phys. Chem. Chem. Phys.* 14, 2250–2254. doi:10.1039/c2cp23499h

Savelyev, A., and Papoian, G. A. (2009). Molecular Renormalization Group Coarse-Graining of Polymer Chains: Application to Double-Stranded DNA. *Biophysical J.* 96, 4044–4052. doi:10.1016/j.bpj.2009.02.067

Schaeffer, J. M., Thachuk, C., and Winfree, E. (2015). "Stochastic Simulation of the Kinetics of Multiple Interacting Nucleic Acid Strands," in *International Workshop on DNA-Based Computers* (Springer), 194–211. doi:10.1007/978-3-319-21999-8_13

Schreck, J. S., Romano, F., Zimmer, M. H., Louis, A. A., and Doye, J. P. (2016). Characterizing DNA star-tile-based Nanostructures Using a Coarse-Grained Model. *ACS nano* 10, 4236–4247. doi:10.1021/acsnano.5b07664

Seeman, N. C. (2003). DNA a Mater. World 421, 427–431. doi:10.1038/nature01406

Sengar, A. (2021). oxDNA Supporting Data. doi:10.5281/zenodo.4809769 (Accessed April 20, 2021).

Seol, Y., Skinner, G. M., Visscher, K., Buhot, A., and Halperin, A. (2007). Stretching of Homopolymeric RNA Reveals Single-Stranded Helices and Base-Stacking. *Phys. Rev. Lett.* 98, 158103. doi:10.1103/PhysRevLett.98.158103

Seol, Y., Skinner, G. M., and Visscher, K. (2004). Elastic Properties of a Single-Stranded Charged Homopolymeric Ribonucleotide. *Phys. Rev. Lett.* 93. doi:10.1103/PhysRevLett.93.118102

Sharma, R., Schreck, J. S., Romano, F., Louis, A. A., and Doye, J. P. (2017). Characterizing the Motion of Jointed DNA Nanostructures Using a Coarse-Grained Model. *ACS nano* 11, 12426–12435. doi:10.1021/acsnano.7b06470

Shi, Z., Castro, C. E., and Arya, G. (2017). Conformational Dynamics of Mechanically Compliant DNA Nanostructures from Coarse-Grained Molecular Dynamics Simulations. *ACS nano* 11, 4617–4630. doi:10.1021/acsnano.7b00242

Shin, J.-S., and Pierce, N. A. (2004). A Synthetic DNA walker for Molecular Transport. *J. Am. Chem. Soc.* 126, 10834–10835. doi:10.1021/ja047543j

Skoruppa, E., Laleman, M., Nomidis, S. K., and Carlon, E. (2017). DNA Elasticity from Coarse-Grained Simulations: The Effect of Groove Asymmetry. *J. Chem. Phys.* 146, 214902. doi:10.1063/1.4984039

Smith, S. B., Cui, Y., and Bustamante, C. (1996). Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science* 271, 795–799. doi:10.1126/science.271.5250.795

Snodin, B. E., Randisi, F., Mosayebi, M., Šulc, P., Schreck, J. S., Romano, F., et al. (2015). Introducing Improved Structural Properties and Salt Dependence into a Coarse-Grained Model of DNA. *J. Chem. Phys.* 142, 234901. doi:10.1063/1.4921957

Snodin, B. E., Romano, F., Rovigatti, L., Ouldridge, T. E., Louis, A. A., and Doye, J. P. K. (2016). Direct Simulation of the Self-Assembly of a Small DNA Origami. *ACS Nano* 10, 1724–1737. doi:10.1021/acsnano.5b05865

Snodin, B. E., Schreck, J. S., Romano, F., Louis, A. A., and Doye, J. P. K. (2019). Coarse-grained Modelling of the Structural Properties of DNA Origami. *Nucleic Acids Res.* 47, 1585–1597. doi:10.1093/nar/gky1304

Šponer, J., Jurečka, P., and Hobza, P. (2004). Accurate Interaction Energies of Hydrogen-Bonded Nucleic Acid Base Pairs. *J. Am. Chem. Soc.* 126, 10142–10151. doi:10.1021/ja048436s

Šponer, J., Riley, K. E., and Hobza, P. (2008). Nature and Magnitude of Aromatic Stacking of Nucleic Acid Bases. *Phys. Chem. Chem. Phys.* 10, 2595–2610. doi:10.1039/b719370j

Srinivas, N., Ouldridge, T. E., Šulc, P., Schaeffer, J., Yurke, B., Louis, A. A., et al. (2013). On the Biophysics and Kinetics of Toehold-Mediated DNA Strand Displacement. *Nucl. Acids Res.* 41, 10641–10658. doi:10.1093/nar/gkt801

Srinivas, N., Parkin, J., Seelig, G., Winfree, E., and Soloveichik, D. (2017). Enzyme-free Nucleic Acid Dynamical Systems. *Science* 358, eaal2052. doi:10.1126/science.aal2052

Stoev, I. D., Cao, T., Caciagli, A., Yu, J., Ness, C., Liu, R., et al. (2020). On the Role of Flexibility in Linker-Mediated DNA Hydrogels. *Soft Matter* 16, 990–1001. doi:10.1039/c9sm01398a

Strick, T. R., Allemand, J. F., and Bensimon, D. (1996). The Elasticity of a Single Supercoiled DNA Molecule (Vol 271, Pg 1835, 1996). *Science* 272, 797. doi:10.1126/science.271.5257.1835

Šulc, P., Romano, F., Ouldridge, T. E., Doye, J. P. K., and Louis, A. A. (2014). A Nucleotide-Level Coarse-Grained Model of RNA. *J. Chem. Phys.* 140, 235102. doi:10.1063/1.4881424

Šulc, P., Romano, F., Ouldridge, T. E., Rovigatti, L., Doye, J. P. K., and Louis, A. A. (2012). Sequence-dependent Thermodynamics of a Coarse-Grained DNA Model. *J. Chem. Phys.* 137, 135101. doi:10.1063/1.4754132

Suma, A., Poppleton, E., Matthies, M., Šulc, P., Romano, F., Louis, A. A., et al. (2019). TacoxDNA: A User-Friendly Web Server for Simulations of Complex DNA Structures, from Single Strands to Origami. *J. Comput. Chem.* 40, 2586–2595. doi:10.1002/jcc.26029

Sutthibutpong, T., Matek, C., Benham, C., Slade, G. G., Noy, A., Laughton, C., et al. (2016). Long-range Correlations in the Mechanics of Small DNA Circles under Topological Stress Revealed by Multi-Scale Simulation. *Nucleic Acids Res.* 44, 9121–9130. doi:10.1093/nar/gkw815

Tee, S. R., and Wang, Z. (2018). How Well Can DNA Rupture DNA? Shearing and Unzipping Forces inside DNA Nanostructures. *ACS omega* 3, 292–301. doi:10.1021/acsomega.7b01692

Tempestini, A., Cassina, V., Brogioli, D., Ziano, R., Erba, S., Giovannoni, R., et al. (2013). Magnetic Tweezers Measurements of the Nanomechanical Stability of DNA against Denaturation at Various Conditions of Ph and Ionic Strength. *Nucleic Acids Res.* 41, 2009–2019. doi:10.1093/nar/gks1206

Tikhomirov, G., Petersen, P., and Qian, L. (2017). Fractal Assembly of Micrometre-Scale DNA Origami Arrays with Arbitrary Patterns. *Nature* 552, 67–71. doi:10.1038/nature24655

Tomov, T. E., Tsukanov, R., Glick, Y., Berger, Y., Liber, M., Avrahami, D., et al. (2017). DNA Bipedal Motor Achieves a Large Number of Steps Due to Operation Using Microfluidics-Based Interface. *Acs Nano* 11, 4002–4008. doi:10.1021/acsnano.7b00547

Torrie, G. M., and Valleau, J. P. (1977). Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comp. Phys.* 23, 187–199. doi:10.1016/0021-9991(77)90121-8

Tortora, M. M., Mishra, G., Prešern, D., and Doye, J. P. (2020). Chiral Shape Fluctuations and the Origin of Chirality in Cholesteric Phases of DNA Origamis. *Sci. Adv.* 6, eaaw8331. doi:10.1126/sciadv.aaw8331

Uusitalo, J. J., Ingolfsson, H. I., Akhshi, P., Tieleman, D. P., and Marrink, S. J. (2015). Martini Coarse-Grained Force Field: Extension to DNA. *J. Chem. Theor. Comput.* 11, 3932–3945. doi:10.1021/acs.jctc.5b00286

Vafabakhsh, R., and Ha, T. (2012). Extreme Bendability of DNA Less Than 100 Base Pairs Long Revealed by Single-Molecule Cyclization. *Science* 337, 1097–1101. doi:10.1126/science.1224139

van Mameren, J., Gross, P., Farge, G., Hooijman, P., Modesti, M., Falkenberg, M., et al. (2009). Unraveling the Structure of DNA during Overstretching by Using Multicolor, Single-Molecule Fluorescence Imaging. *Proc. Natl. Acad. Sci.* 106, 18231–18236. doi:10.1073/pnas.0904322106

Wagenbauer, K. F., Sigl, C., and Dietz, H. (2017). Gigadalton-scale Shape-Programmable DNA Assemblies. *Nature* 552, 78–83. doi:10.1038/nature24651

Wang, M. D., Yin, H., Landick, R., Gelles, J., and Block, S. M. (1997). Stretching DNA with Optical Tweezers. *Biophysical J.* 72, 1335–1346. doi:10.1016/s0006-3495(97)78780-0

Wang, Q., Irobalieva, R. N., Chiu, W., Schmid, M. F., Fogg, J. M., Zechiedrich, L., et al. (2017). Influence of DNA Sequence on the Structure of Minicircles under Torsional Stress. *Nucleic Acids Res.* 45, 7633–7642. doi:10.1093/nar/gkx516

Wang, Q., Myers, C. G., and Pettitt, B. M. (2015). Twist-induced Defects of the P-SSP7 Genome Revealed by Modeling the Cryo-EM Density. *J. Phys. Chem. B* 119, 4937–4943. doi:10.1021/acs.jpcb.5b00865

Wang, Q., and Pettitt, B. M. (2014). Modeling DNA Thermodynamics under Torsional Stress. *Biophysical J.* 106, 1182–1193. doi:10.1016/j.bpj.2014.01.022

Wang, Q., and Pettitt, B. M. (2016). Sequence Affects the Cyclization of DNA Minicircles. *J. Phys. Chem. Lett.* 7, 1042–1046. doi:10.1021/acs.jpclett.6b00246

Wang, X., and Ha, T. (2013). Defining Single Molecular Forces Required to Activate Integrin and Notch Signaling. *Science* 340, 991–994. doi:10.1126/science.1231041

Watson, J. D., and Crick, F. H. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737–738. doi:10.1038/171737a0

Whitelam, S., Feng, E. H., Hagan, M. F., and Geissler, P. L. (2009). The Role of Collective Motion in Examples of Coarsening and Self-Assembly. *Soft Matter* 5, 1251–1262. doi:10.1039/b810031d

Whitelam, S., and Geissler, P. L. (2007). Avoiding Unphysical Kinetic Traps in Monte Carlo Simulations of Strongly Attractive Particles. *J. Chem. Phys.* 127, 154101. doi:10.1063/1.2790421

Wickham, S. F. J., Bath, J., Katsuda, Y., Endo, M., Hidaka, K., Sugiyama, H., et al. (2012). A DNA-Based Molecular Motor that Can Navigate a Network of Tracks. *Nat. Nanotechnol.* 7, 169–173. doi:10.1038/nnano.2011.253

Widom, J. (2001). Role of DNA Sequence in Nucleosome Stability and Dynamics. *Q. Rev. Biophys.* 34, 269. doi:10.1017/s0033583501003699

Woods, D., Doty, D., Myhrvold, C., Hui, J., Zhou, F., Yin, P., et al. (2019). Diverse and Robust Molecular Algorithms Using Reprogrammable DNA Self-Assembly. *Nature* 567, 366–372. doi:10.1038/s41586-019-1014-9

Xayaphoummine, A., Bucher, T., and Isambert, H. (2005). Kinefold Web Server for RNA/DNA Folding Path and Structure Prediction Including Pseudoknots and Knots. *Nucleic Acids Res.* 33, W605–W610. doi:10.1093/nar/gki447

Yamakawa, H. (1977). A Hypothesis on Polymer Chain Configurations. Helical Wormlike Chains. *Macromolecules* 10, 692–696. doi:10.1021/ma60057a039

Yao, G., Zhang, F., Wang, F., Peng, T., Liu, H., Poppleton, E., et al. (2020). Meta-dna Structures. *Nat. Chem.* 12, 1067–1075. doi:10.1038/s41557-020-0539-8

Yoo, J., and Aksimentiev, A. (2012). Improved Parametrization of Li+, Na+, K+, and Mg2+ Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. *J. Phys. Chem. Lett.* 3, 45–50. doi:10.1021/jz201501a

Yurke, B., Turberfield, A. J., Mills, A. P., Simmel, F. C., and Neumann, J. L. (2000). A DNA-Fuelled Molecular Machine Made of DNA. *Nature* 406, 605–608. doi:10.1038/35020524

Zhang, D. Y., and Seelig, G. (2011). Dynamic DNA Nanotechnology Using Strand-Displacement Reactions. *Nat. Chem.* 3, 103–113. doi:10.1038/nchem.957

Zhang, F., Jiang, S., Wu, S., Li, Y., Mao, C., Liu, Y., et al. (2015). Complex Wireframe DNA Origami Nanostructures with Multi-Arm junction Vertices. *Nat. Nanotechnol.* 10, 779. doi:10.1038/nnano.2015.162

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.