A nucleotide-level coarse-grained model of RNA

Petr Šulc, Flavio Romano, Thomas E. Ouldridge, Jonathan P. K. Doye, and Ard A. Louis

Articles you may be interested in

Coarse-grained modelling of supercoiled RNA
J. Chem. Phys. **143**, 243122 (2015); 10.1063/1.4933066

Introducing improved structural properties and salt dependence into a coarse-grained model of DNA
J. Chem. Phys. **142**, 234901 (2015); 10.1063/1.4921957

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

A coarse-grain three-site-per-nucleotide model for DNA with explicit ions
J. Chem. Phys. **135**, 165104 (2011); 10.1063/1.3652956

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

A nucleotide-level coarse-grained model of RNA

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

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

We present a new, nucleotide-level model for RNA, oxRNA, based on the coarse-graining methodology recently developed for the oxDNA model of DNA. The model is designed to reproduce structural, mechanical, and thermodynamic properties of RNA, and the coarse-graining level aims to retain the relevant physics for RNA hybridization and the structure of single- and double-stranded RNA. In order to explore its strengths and weaknesses, we test the model in a range of nanotechnological and biological settings. Applications explored include the folding thermodynamics of a pseudoknot, the formation of a kissing loop complex, the structure of a hexagonal RNA nanoring, and the unzipping of a hairpin motif. We argue that the model can be used for efficient simulations of the structure of systems with thousands of base pairs, and for the assembly of systems of up to hundreds of base pairs. The source code implementing the model is released for public use. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4881424]

I. INTRODUCTION

RNA (ribonucleic acid) strands are polymers that play crucial cellular roles in gene expression, translation and regulation.[1] RNA is composed of units called nucleotides, which consist of a phosphate and ribose sugar backbone to which one of four different bases is attached: adenine (A), guanine (G), cytosine (C) or uracil (U). DNA has a similar chemical structure, but thymine (T) is present in place of uracil and the backbone includes deoxyribose sugars instead of ribose sugars. Both RNA and DNA can form double-helical molecules, stabilized by hydrogen bonds between complementary Watson-Crick base pairs: AU (AT in case of DNA) and GC. For RNA, wobble base pairs (GU) can also stabilize the duplex form and are commonly found. While DNA typically forms a B-form helix, double-stranded RNA is typically found in a related, but different A-form helical geometry.[2]

DNA's role in biological systems is to store genetic information and it is most often found in its double-helical form. By contrast, most naturally occurring RNA molecules are single-stranded and fold into complex structures that contain double-helical segments as well as loops, junctions, and numerous tertiary structure interactions that further stabilize the molecule. The lengths of RNA strands found in living cells can range from a few nucleotides to several thousands or even tens of thousands.[3]

DNA nanotechnology[4] uses DNA molecules to create nanoscale structures and active devices. Examples of experimentally realized systems include DNA motors,[5,6] self-assembled nanostructures such as DNA origamis[7] and the use of DNA strands for computation.[8] Since RNA molecules are more difficult to handle and preserve than DNA in experimental settings, developing RNA-based nanodevices has been a more challenging task. However, the emerging field of RNA nanotechnology[9] offers promising applications of RNA-based nanodevices both *in vitro* and *in vivo*. The general design principles from DNA nanotechnology can be applied to RNA, but because specific RNA sequences form functional structures that are known to interact with proteins and other RNA molecules in the cell, RNA molecules would be the natural choice for nanotechnology devices operating inside cells.[10] Successful examples of experimentally realized RNA nanotechnology include self-assembling RNA nanocubes[11] and RNA tiles.[12] Recently, a RNA strand displacement reaction cascade was proposed that could be used for conditional gene silencing inside the cell.[13]

Because of its importance for biological systems, RNA has been the subject of intensive study. In addition to numerous experimental studies, a range of theoretical and computational methods have been applied to the study of its properties. Currently, there are multiple tools and computational approaches that can describe RNA structure at various levels of detail.[14–16]

In an important series of works[17–23] the thermodynamics of RNA secondary structure (i.e., the list of base pairs in the folded state of RNA) was characterized in terms of a nearest-neighbor model for calculating the free energies of various secondary-structure motifs. The nearest-neighbor model is the basis of various tools for the prediction of the secondary structure.[24–26] Furthermore, some tools have been extended by adding simple kinetic descriptions to the nearest-neighbor thermodynamics, allowing folding transitions to be modeled.[27,28] Although these methods are typically very fast, the fundamentally discrete nature of the description and the lack of structural and mechanical detail places a limit on what they can treat. Similar caveats apply to other models of RNA base pairing with reduced dimensionality.[29]

At the highest level of detail, quantum chemistry methods have been employed to study the interactions between nucleotides.[30,31] Due to the complexity of the calculations, such methods remain limited to interactions between nearest-neighbor base pairs in vacuum. Fully atomistic molecular dynamics simulation packages such as AMBER[32] or CHARMM[33] include the solvent, and use classical effective interaction potentials between atoms to represent systems with high resolution. However, the study of rare events such as the formation or breaking of individual base pairs remains a very challenging task. Simulations of several folding pathways of a short RNA hairpin[34] and tetraloops[35] provide examples of the limit of what is currently possible. At present, atomistic molecular dynamics simulations can attain time scales of the order of μs. Methods that combine fully atomistic representation with hierarchical Monte Carlo sampling were used for conformational sampling of larger structures such as tRNAs.[36,37] Moreover, while the forcefields are improving, they are still under development so that different versions can generate different behavior.[38–40]

In order to access longer time scales relevant to rare events, such as the breaking of base pairs or the formation of large structures, one needs to use a more coarse-grained description. In this approach, atoms are incorporated into a reduced set of degrees of freedom that experience effective interactions. Solvent molecules are often integrated out. Such models always present a compromise between accuracy, efficiency and the level of detail, which determines their scope. Several coarse-grained RNA models have been developed in recent years.

Most coarse-grained models are primarily designed to predict folded structures of RNA sequences.[41–45] These models typically use potentials obtained from analyses of crystal structures,[41,45–47] homology modelling[42,43] or microscopic atomistic simulations.[44] The data to which these models were parameterized do not reflect the thermodynamics of RNA self-assembly, and are biased towards the properties of folded conformations. It is therefore difficult to use these models to describe actual assembly processes and reactions.

The models of Ding *et al.*[48] and Denesyuk and Thirumalai[49] choose interaction strengths based on the nearest-neighbor model of thermodynamics. Indeed, the model of Denesyuk and Thirumalai (which is an extension of Ref. 50) has been used to study folding thermodynamics in comparison with experiment for hairpins and a pseudoknot. The thermodynamic behavior of these models has not, however, been tested against a wide range of systems and detailed investigations of their mechanical properties are yet to be performed.

Lattice models of RNA have been designed with the explicit goal of reproducing folding thermodynamics.[51,52] While these models can be efficient and thermodynamically accurate, they are limited in scope by their lattice-based description.

We note that mechanical properties have not been reported for any of these coarse-grained models. If thermodynamic properties were considered, they were usually studied for only few selected systems.

Here we propose a new off-lattice coarse-grained RNA model, oxRNA, that follows the coarse-graining approach developed for the DNA model oxDNA.[53–55] Given that oxDNA has been successfully used to model DNA nanotechnological systems, such as motors,[56,57] tweezers,[58] kissing hairpins,[53,59] strand displacement,[60] as well as for biophysical applications including cruciforms,[61] the pulling of single-[53] and double-stranded DNA,[62] and the hybridization of short oligonucleotides,[63] our goal is to derive a model of similar applicability for RNA. We aim to capture average RNA structure and mechanics, and sequence-dependent RNA thermodynamics, in as minimalistic a description as possible. We replace each RNA nucleotide by a single rigid body with multiple interaction sites. The interactions between rigid bodies are parametrized to allow an A-form helix to form from two single strands and to reproduce RNA thermodynamics. The resultant model goes beyond nearest-neighbor thermodynamics because it has the ability to capture topological, mechanical, and spatial effects and allows for the study of kinetic properties of various processes within a molecular dynamics simulation framework. The model uses only pairwise interactions to facilitate the use of cluster Monte Carlo algorithms for simulations. The simple representation, one rigid body per nucleotide, allows for efficient simulation of structures of sizes up to several hundred nucleotides on a single CPU as well as of rare events such as the dissociation or the formation of a double helix.

This paper is organized as follows. In Sec. II we present our coarse-grained model and its parametrization. In Sec. III we study the thermodynamic, structural, and mechanical properties of the model. We then illustrate the utility of the model through applications to pseudoknot thermodynamics, hairpin unzipping, and kissing hairpins in Sec. IV. The detailed description of the interaction potentials in our model is provided in the supplementary material.[64]

II. THE RNA MODEL AND ITS PARAMETRIZATION

In this section, we describe the parametrization of the oxRNA model to reproduce the RNA thermodynamics of the nearest-neighbor model, which we briefly outline in Sec. II A.

A. RNA thermodynamics and the nearest-neighbor model

In an extensive series of investigations,[18–20,23] Turner *et al.* parametrized a nearest-neighbor model (hereafter referred to as the NN-model) to describe the thermodynamics of RNA duplex and hairpin formation, which is widely used in RNA secondary structure prediction.[24,28,65–67] The model treats RNA at the level of secondary structure, estimating enthalpic and entropic contributions to the stability from each pair of consecutive base pairs (bp) in a structure and including corrections for end effects and enclosed loops of unpaired bases. The parametrization used melting experiments of short duplexes and hairpins at 1 M [Na^+]. The results were fitted using a two-state assumption in which RNA either adopts the fully-formed structure or is completely disordered. The yield of the duplexes is then given by

$$\frac{[AB]}{[A][B]} = \exp(-\Delta G^{\ominus}(T)/RT), \qquad (1)$$

where $[AB]$ is the molar concentration of the duplexes, and $[A]$ and $[B]$ the concentrations of the single strands. $\Delta G^{\ominus}(T) = \Delta H^{\ominus} - T\Delta S^{\ominus}$ is the standard Gibbs free-energy change that is given by the NN-model, where $\Delta H^{\ominus}$ and $\Delta S^{\ominus}$ are calculated from the base-pair step contributions. The concentrations of reactants are specified relative to 1M and $\Delta G^{\ominus}$ is calculated for a system where all reactants have a concentration of 1M.

Similarly, the yields of hairpins are

$$\frac{[C]}{[O]} = \exp(-\Delta G^{\ominus}(T)/RT), \quad (2)$$

where $[C]$ is the concentration of closed strands, and $[O]$ is the concentration of open strands.

The melting temperature of a duplex, at a given strand concentration, is defined as the temperature at which half of the duplexes present in the solution are dissociated into single strands. Similarly, the melting temperature of a hairpin is defined as the temperature at which the strand has a 50% probability of being open.

The NN-model has been shown to reproduce the melting temperatures of RNA oligonucleotides with Watson-Crick base-pairing with 1.3 °C accuracy.[18] In our work, we will treat the NN-model as an accurate fit to the melting data and use its melting temperatures predictions for fitting the oxRNA model. The average accuracy (the fraction of correctly predicted base pairs in known secondary structures) of the NN-model is reported to be 73% for a database of strands that has in total 150 000 nucleotides, containing domains of up to 700 nucleotides.[19]

B. The representation

OxRNA uses a single rigid body with multiple interaction sites to represent a nucleotide. The rigid bodies have the same size and shape for all nucleotides. Each rigid body has a backbone, 3′-stacking, 5′-stacking, cross-stacking, and hydrogen-bonding interaction sites. The detailed description of the representation of a nucleotide is given in the supplementary material (Fig. S2).[64] In the figures we use a schematic ellipsoid to represent the stacking and hydrogen-bonding sites as this allows the orientation of the base to be clearly seen. The potentials between the nucleotides are effective interactions that are designed to capture the overall thermodynamic and structural consequence of the base-pairing and stacking interactions, rather than directly representing the microscopic contributions such as electrostatics, dispersion, exchange repulsion, and hydrophobicity.

We choose the functional forms of our coarse-grained interactions to reproduce directly experimentally measured properties of RNA. For these reasons, we label our coarse-graining approach "top-down," as opposed to a "bottom-up" approach which starts from a more detailed description of the system. We point out that any coarse-grained interaction is actually a free energy for the real system, rather than a potential energy, and therefore it is in principle state-point dependent. So it should come to no surprise that our potential contains an explicit dependence on the temperature, although for simplicity we try to limit this as much as possible and only introduce

it in one of the interaction terms (V_{stack}, as we will discuss later). Our coarse-graining aims to retain the relevant geometric degrees of freedom in order to still correctly capture the relative entropies of different states, despite not having temperature dependence in most of the interaction potentials.[68]

The potential energy of the model is

$$V_{oxRNA} = \sum_{\langle ij \rangle} (V_{backbone} + V_{stack} + V'_{exc})$$
$$+ \sum_{i,\,j \notin \langle ij \rangle} (V_{H.B.} + V_{cross\,st.} + V_{exc} + V_{coaxial\,st.}), \quad (3)$$

where the first sum is taken over all the nucleotides that are neighbors along a RNA strand and the second sum is taken over all the non-nearest-neighbor pairs of nucleotides. All potentials are two-body potentials. There is a maximum distance beyond which all potentials are zero (with the exception of $V_{backbone}$ which diverges to infinity as the distance between adjacent backbone sites approaches its maximum value). The interactions are schematically shown in Fig. 1. We discuss briefly the potentials here while the detailed description is given in Sec. S-III B of the supplementary material.[64]

We note that the positions of the minima in the potential functions have been selected so that the model reproduces the structure of the A-RNA double helix, which RNA duplexes have been shown to adopt[1,2] and which we describe in more detail in Sec. III A. The widths of the potential functions and the strengths of the interaction potentials were parametrized to reproduce RNA thermodynamics as described in Sec. II D. Since all nucleotides in oxRNA have the same size and shape, and the position of the energy minima for the potentials are the same for all types of nucleotides, we do not expect oxRNA to capture detailed sequence-dependent structure of the A-form helix. We aim to capture sequence-dependent thermodynamics by setting the strength of some of the interactions to be sequence dependent, as was the case also for oxDNA. Although in principle this variety of interaction strengths could lead to sequence-dependent mechanical properties, we
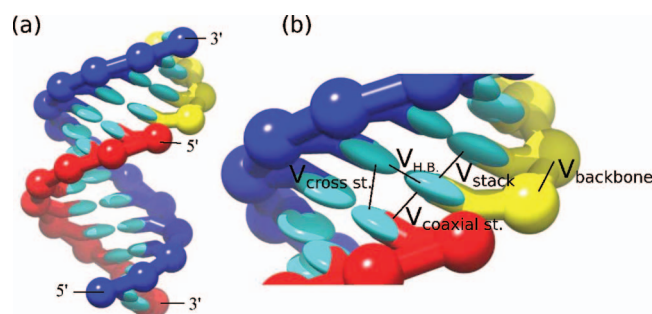

FIG. 1. A schematic representation of (a) an A-RNA helix as represented by the model and of (b) the attractive interaction in oxRNA. The lines in (b) schematically show the interactions between the nucleotides: Hydrogen bonding ($V_{H.B.}$) between complementary base pairs on opposite strands, stacking (V_{stack}) between adjacent nucleotides on the same strand, cross-stacking ($V_{cross\,st.}$) between a nucleotide and the nucleotide that is the 3′ neighbor of the directly opposite nucleotide, coaxial stacking ($V_{coaxial\,st.}$) that is a stacking interaction between nucleotides that are not neighbors on the same strand, and backbone interaction ($V_{backbone}$) that constrains the distance between neighboring nucleotides. The nucleotides can also interact with excluded-volume interactions, which are not shown.

have not optimized the model in this regard. We do this for simplicity, and the thermodynamic consequences of sequence dependence are stronger and better characterized than the mechanical effects.

The backbone interaction, V_{backbone}, is an isotropic FENE (finitely extensible nonlinear elastic) potential and depends only on the distance between the backbone sites of the two adjacent nucleotides. This potential is used to mimic the covalent bonds in the RNA backbone that constrain this intramolecular distance. The nucleotides also have repulsive excluded-volume interactions V_{exc} and V'_{exc} that depend on the distance between the interaction sites, namely the backbone-backbone, stacking-stacking, and stacking-backbone distances. The excluded-volume interactions ensure that strands cannot overlap, or pass through each other in a dynamical simulation.

The duplex is stabilized by hydrogen bonding ($V_{\text{H.B.}}$), stacking (V_{stack}), and cross-stacking ($V_{\text{cross st.}}$) interactions. These potentials are highly anisotropic and depend on the distance between the relevant interaction sites as well as the mutual orientations of the nucleotides. The anisotropic potentials are of the form

$$V_{\text{H.B.}} = \alpha_{ij} f_{\text{H.B.}}(\mathbf{r}_{ij}, \mathbf{\Omega}_i, \mathbf{\Omega}_j), \tag{4}$$

$$V_{\text{stack}} = \eta_{ij}(1 + \kappa \, k_B T) f_{\text{stack}}(\mathbf{r}_{ij}, \mathbf{\Omega}_i, \mathbf{\Omega}_j), \tag{5}$$

$$V_{\text{cross st.}} = \gamma f_{\text{cross st.}}(\mathbf{r}_{ij}, \mathbf{\Omega}_i, \mathbf{\Omega}_j), \tag{6}$$

$$V_{\text{coaxial st.}} = \mu f_{\text{coaxial st.}}(\mathbf{r}_{ij}, \mathbf{\Omega}_i, \mathbf{\Omega}_j), \tag{7}$$

where the functions f are a product of multiple terms, one of which depends on the distance between the relevant interaction sites and the remaining are angular modulation functions that are equal to one if the relevant angles between the nucleotides correspond to the minimum potential energy configuration, and smoothly go to zero as they depart from these values. The set of angles is different for each potential and includes angles between intersite vectors and orientations $\mathbf{\Omega}_i$ and $\mathbf{\Omega}_j$ of the nucleotides. The constant prefactors α_{ij}, η_{ij}, γ, and μ set the strength of the interactions, with α_{ij}, η_{ij} being dependent on the nucleotides involved. We note that while it is possible to make γ and μ also sequence dependent, it was found during the parametrization that the thermodynamic predictions of the model do not improve with the sequence-dependent $V_{\text{cross st.}}$ and $V_{\text{coaxial st.}}$ and therefore their interaction strengths were set to be the same for all nucleotides.

The hydrogen-bonding term $V_{\text{H.B.}}$ is designed to capture the duplex stabilizing interactions between Watson-Crick and wobble base pairs. The potential reaches its minimum when two complementary nucleotides (AU, GC or GU) are coplanar, directly opposite and antiparallel with respect to each other and at the right distance.

The stacking interaction V_{stack} mimics the favorable interaction between adjacent bases, which results from a combination of hydrophobic, electrostatic and dispersion effects. It acts only between nearest-neighbor nucleotides and its strength depends on both the distance between the respective 3′ and 5′ stacking sites of the nucleotides as well as their

mutual orientations. It also depends on the vector between the backbone interaction sites in a way that ensures the inclination of the bases in the duplex structure matches that for A-RNA. We note that the nucleotides can also interact via the stacking interaction when they are in the single-stranded state. To ensure the right-handedness of the RNA helix in the duplex state, the stacking interaction has an additional modulation term that is equal to one if the nucleotides adopt a right-handed conformation and goes smoothly to zero in the left-handed conformation.

Similarly to oxDNA, the interaction strength of the stacking potential has a temperature-dependent contribution (the term $\kappa \, k_B T$ in Eq. (5)). This term was introduced in oxDNA in order to correctly reproduce the thermodynamics of the stacking transition. We also found that retaining this temperature dependence enables oxRNA to reproduce more accurately the widths of the melting transitions, which are discussed in more detail in Sec. II D.

The cross-stacking potential, $V_{\text{cross st.}}$, is designed to capture the interactions between diagonally opposite bases in a duplex and has its minimum when the distance and mutual orientation between nucleotides correspond to the arrangement of a nucleotide and a 3′ neighbor of the directly opposite nucleotide in a A-form helix. This interaction has been parametrized to capture the stabilization of a RNA duplex by a 3′ overhang.[20] OxRNA does not include any interaction with the 5′ neighbor of the directly opposite nucleotide, as 5′ overhangs are significantly less stabilizing than 3′ overhangs.[20]

Finally, the coaxial stacking potential $V_{\text{coaxial st.}}$ represents the stacking interaction between nucleotides that are not nearest-neighbors on the same strand.

Although our model does not include an explicit term for electrostatic interactions between phosphates, these interactions are effectively incorporated into the backbone repulsion. We chose to parametrize our model to the experimental data at 1 M [Na+], where the electrostatic interaction is highly screened, making our approach reasonable. Furthermore, we are only able to capture those tertiary structure motifs that involve Watson-Crick and wobble base pairing or stacking, such as kissing hairpins or coaxial stacking of helices. In particular, oxRNA does not include Hoogsteen or sugar-edge hydrogen-bonded base pairs, or ribose zippers (interactions involving the 2′-OH group on the ribose sugar). In principle, these interactions could be included, but for this version of the model we chose not to as there are no systematic thermodynamic data to which we could parametrize the relevant interactions.

C. Simulation methods

1. Algorithms

For the majority of our simulations, unless noted otherwise, we use the Virtual Move Monte Carlo algorithm (VMMC) (specifically the variant described in the Appendix of Ref. 69) to obtain the thermodynamics of oxRNA. VMMC is a cluster-move Monte Carlo algorithm that has previously been used in many applications for the oxDNA model.[68] We also implemented the oxRNA force field in a molecular dynamics (MD) code with the choice of a Langevin[70] or an

Andersen-like[71] thermostat. For the MD simulations in this work, we used the Andersen-like thermostat.

When calculating the free energy of a system as a function of an order parameter (typically the number of base pairs between two strands or, for example, the number of base pairs in a hairpin stem), we run multiple simulations that sample all relevant states. We then calculate p_i, the probabilities of the system having a particular value i of the order parameter, and take the free energy of such a state to be $F_i = -k_B T \ln(p_i) + c$, where k_B is the Boltzmann constant and c is a constant. We are interested only in differences in free energies between different states and hence c can be set to an arbitrary value.

In our sampling simulations, we often combine the VMMC algorithm with the umbrella sampling method[72] in order to help systems overcome free-energy barriers, a technique which we briefly discuss in Sec. S-II of the supplementary material.[64]

2. Calculation of melting temperatures

When calculating the melting temperature of a duplex, we use VMMC with umbrella sampling to simulate a system of two complementary strands and calculate the ratio Φ of the times that the system spends in a duplex state and in a dissociated state. At the melting temperature T_m, $\Phi = 2$ for heterodimers, accounting for finite-size effects that come from simulating only two strands, as explained in Refs. 73–75. For transitions involving single strands, such as hairpin formation, the ratio Φ is the time spent in closed states (i.e., states with at least one base pair in the stem) divided by the time spent in open states. In this case, $\Phi = 1$ at the melting temperature because the finite size effects mentioned above are irrelevant for unimolecular assembly. We note that we consider the system to be in a duplex (hairpin) form if the complementary strands (stems) have one or more base pairs present.

Using the histogram reweighting method, the states generated during the umbrella sampling simulations can be used to calculate the melting temperatures for a model with the same functional forms of the potential but with different interaction strengths. Previously, we used this approach for the parametrization of the oxDNA model and the method is described in detail in Ref. 53. The reweighting method recalculates the ratio Φ for temperature T and a given set of interaction strengths by counting the contribution to the bonded or unbonded state of each sampled state with an additional weight factor

$$w = \exp\left(\frac{V_0(T_0)}{k_B T_0} - \frac{V(T)}{k_B T}\right),$$

where V_0 is the potential energy of the state generated in the simulation with the original set of parameters at temperature T_0, and $V(T)$ is the potential energy of the same state recalculated for the interaction strengths for which we want to find the ratio Φ at temperature T. The reweighting method allows us to extract Φ for a given set of interaction strengths at a range of temperatures which then can be interpolated to find the melting temperature. This method assumes that the ensemble of states generated with the original set of parameters is representative of the states that the system would visit in a simulation with the new parameters. Using this approach it is possible to calculate the melting temperatures for thousands of different sets of interaction constants in an hour of CPU time without the necessity of rerunning the umbrella sampling simulation, which, by contrast, can take several days on a single CPU for each sequence to calculate the melting temperature to within $1\,^\circ$C accuracy.

D. Parametrization of the model

The anisotropic potentials V_{stack}, $V_{\text{H.B.}}$, and $V_{\text{cross st.}}$ have interaction strengths of the form $\eta_{ij}(1 + \kappa k_B T)$, α_{ij}, and γ, respectively, where the stacking interaction strength depends also on the simulation temperature T and i and j correspond to the types of interacting nucleotides (A, C, G, U). The magnitude of the temperature dependence of the stacking, κ, and the cross stacking interaction strength γ are set to be the same for all nucleotide types.

In the first step of the fitting procedure, we parametrize the model to reproduce the melting temperatures of the averaged NN-model, for which we define the enthalpy and entropy contribution per base-pair step by averaging contributions of all possible Watson-Crick base-pair steps in the NN-model. In calculating average melting temperatures of different motifs, such as hairpins or terminal mismatches and internal mismatches, the additional entropy and enthalpy contributions for a particular motif in the NN-model were again averaged over all possible combinations of bases. In the averaged NN-model, the melting temperature is hence independent of the particular sequence, but depends only on the lengths of the sequence and the particular secondary structure motif.

The fitting of the interaction strength parameters was done by a simulated annealing algorithm, which aims to find a set of parameters that minimizes the sum of absolute differences between the melting temperatures of a set of systems as calculated by oxRNA and as predicted by the NN-model. The algorithm uses the reweighting method outlined in Sec. II C 2 on an ensemble of states generated by a VMMC simulation to calculate melting temperatures for a particular set of parameters. This algorithm for model parametrization is described in detail in Ref. 53.

First, the oxRNA model was parametrized to reproduce averaged NN-model melting temperatures of structures with only Watson-Crick base pairs. The interaction strength η_{ij} was hence set to η_{avg} for all base pair types i and j and α_{ij} was set to α_{avg} for Watson-Crick complementary nucleotides and 0 otherwise. The initial values for α_{avg}, η_{avg}, and γ were first chosen by hand and then refined based on results of VMMC simulations in order to reproduce melting temperatures as predicted by the averaged NN-model of short duplexes of lengths 5, 6, 7, 8, 10, and 12 bp and of duplexes of lengths 5', 6, 8 bp with one overhanging nucleotide at either both 3' ends or both 5' ends. We set κ to be equal to 1.9756 (in the inverse of the energy unit used by our simulation code, as defined in Sec. S-III of the supplementary material[64]), the same value as was used by the previous oxDNA model.[53,54] We found that leaving κ as a free fitting parameter did not lead to a significantly better fit to the considered sequences and motifs.

We note that for some applications, where one is more interested in the qualitative nature of the studied system or one wants to average over all possible sequences, it might be more useful to study the system with a sequence independent model. We refer to such a model as the "average-base" model meaning that η_{ij} are set to η_{avg} for all types of bases and α_{ij} are set to α_{avg} for all Watson-Crick complementary base pairs (GC and AU) and 0 otherwise. If one is interested in sequence-specific effects, then a more complex parametrization is necessary.

We used the final values of η_{avg}, α_{avg} as the initial values for fitting the sequence-dependent values η_{ij}, α_{ij}, with i and j being Watson-Crick or wobble base pairs (AU, GC, GU). The parameters were fitted to an ensemble that contained oligomers of the above mentioned sizes and hairpins with stem lengths 6, 8 and 10 and loop lengths 5, 6, 7, 8, 10 and 15, with 100 randomly generated sequences (with only Watson-Crick base pairs) for each size, and 533 further random duplexes of lengths 5–12 bp containing GU wobble base pairs. We excluded sequences with neighboring wobble base pairs in the fitting process as these can lead to duplexes with particularly low melting temperatures (some of the base pair steps containing wobble base pairs are actually destabilizing at room temperature[20]). We found that our model was unable to accurately fit melting temperatures of duplexes that contain neighboring GU/UG or UG/GU wobble base pairs, probably due to the fact that we do not account for the structural changes that these induce in the duplex.

We note that if one included only Watson-Crick base pairs in the sequence-dependent fitting (as was the case for the parametrization of the oxDNA model[53]), it would not be possible to distinguish between certain stacking interaction types. For instance, the contribution of AA and UU base stacking interactions always appear together in the AA/UU base pair step free-energy contribution in the NN-model. However, including wobble base pairs in the fitting ensemble provides additional information, for example, the UU stacking contribution also appears in the AG/UU base-pair step. We therefore do not need to restrict the strength of stacking interaction to be the same for certain types of nucleotides, as was the case for the oxDNA model.

Finally, we parametrized the coaxial stacking interaction potential, $V_{\mathrm{coaxial\ st.}}$, which captures the stacking interaction between two bases that are not neighbors along the same strand. Experiments have measured this interaction by a comparison of the melting of a 4-base strand with its complement, or with a hairpin with a 4-base overhang with the complementary sequence adjacent to the hairpin stem. They found for both DNA and RNA that the melting temperature increases for the 4bp long strand attached to the overhang on the hairpin stem, which was attributed to the extra stabilizing interactions with the adjacent stem.[21,22,76,77] The coaxial stacking free energy has been incorporated in the NN-model by assuming that the free-energy stabilizations in these experiments are similar in strength to the actual base-pair steps with the same sequence. The NN-model hence uses the same free energy contribution for a base pair coaxially stacked on a subsequent base pair (as illustrated in Fig. 1) as it uses for a base pair step in an uninterrupted duplex. In order to parametrize these in-

teractions for oxRNA, we performed melting simulations of a 5-base strand, which was able to associate with a complementary 5′ overhang on a longer duplex (which itself was stable). We fitted the interaction strength μ of the coaxial stacking interaction in our model so that it would match the prediction of the melting temperature by the averaged NN-model. We note that in our model, the contributing factors to stabilization are both the coaxial stacking interaction and the cross-stacking interaction between the 5-base strand and the hairpin.

III. PROPERTIES OF THE MODEL

In this section, we describe the structural properties of the model and report the thermodynamics of duplexes, hairpins, and other secondary structure motifs as represented by the model. We further study some of its mechanical properties, namely, the persistence length of a duplex, the force-extension curve for duplex stretching, and the overstretching transition.

A. Structure of the model

As mentioned in Sec. II D, the coarse-grained interactions were selected so that the model reproduces the A-form helix that RNA duplexes have been shown to adopt at physiological conditions.[2,78]

The A-RNA structure is significantly different from B-DNA, the prevalent duplex structure found in DNA molecules. These differences are mainly caused by the sugars in A-RNA adopting a more twisted conformation (C3′ *endo* pucker) as a result of the presence of an extra OH group on the sugar. The A-RNA duplex has a reported helical twist ranging[78] from 32.7° to 33.5° per base pair, corresponding to a pitch of 10.7 to 11 base pairs. The rise per base pair reported by X-ray measurements[78] is about 0.28 nm. The bases are displaced from the helical axis, i.e., the helical axis does not pass through the base pair midpoints as it is approximately the case for the B-DNA helix. Finally, the bases are not perpendicular to the helical axis, but are inclined at an angle of about 15.5°. Although the width of A-RNA is reported to be about 2.1 nm from X-ray crystal structures,[79] Kebbekus *et al.*[80] uses an effective hydrodynamic diameter of 2.8 nm for the structure. The A-RNA helix has a narrow major groove (0.47 nm) and a wide minor groove (1.08 nm).

To characterize the structure of the oxRNA duplex, we simulate a 13-base-pair duplex at 25 °C using Monte Carlo. We generated 30 000 decorrelated configurations that were analyzed in the following manner. The helical axis was fitted for each saved configuration, as described in Sec. S-I of the supplementary material.[64] The rise per base pair was measured as the distance between the projections of the midpoints of base pairs onto the helical axis. The length scale in the oxRNA model is defined so that the rise per base pair is 0.28 nm. The twist per base pair was measured as the angle between the projections of the vectors connecting bases in the base pairs onto the plane perpendicular to the helical axis. The mean turn per base pair in the model is 33.0°, corresponding to a pitch of 10.9 base pairs. The measurements are schematically illustrated in Sec. S-I of the supplementary material

(Fig. S1).[64] The inclination, measured as the mean angle between the vector pointing from the center of mass of a nucleotide to its base and the plane perpendicular to the helical axis, is 16.1°.

The width of the helix is measured as twice the distance of the backbone from the axis, and includes the excluded volume interaction radius of the backbone site. The helix width in oxRNA is 2.5 nm. The major and minor grooves in oxRNA are 0.48 nm and 1.07 nm, respectively, where we measured the groove distances in a manner analogous to a method employed by the Curves+ software[81] for analyzing atomistic structures of DNA and RNA. For a selected nucleotide, we measured distances between its backbone site and points on a curve that was linearly interpolated through the backbone sites of the nucleotides on the opposite strand. The distances measured along the curve have two minima, one for each groove. The excluded volume interaction radius for each backbone site was subtracted from these measured distances.

B. Thermodynamics of the model

In this section, we examine the thermodynamics of duplexes, hairpins, bulges, and internal and terminal mismatches as represented by oxRNA. We compare the melting temperatures as predicted by oxRNA with the melting temperatures calculated from the NN-model (denoted as $T_m(NN)$) for different sequences and different secondary structure motifs. To calculate the melting temperatures, we used the reweighting method, described in Sec. II C 2, with the states that were generated from VMMC simulations of melting for the average-base parametrization of oxRNA.

1. Duplex and hairpin melting

A comparison of the melting temperatures of the average-base parametrization of oxRNA with the thermodynamics of the averaged NN-model for structures involving only Watson-Crick base pairs is shown in Table I. For this averaged model,

TABLE I. The melting temperatures of a series of duplexes for the average-base parametrization of oxRNA (T_m) compared to the averaged NN-model ($T_m(NN^{avg})$). The melting temperatures were calculated from VMMC simulations and are for a strand concentration of 3.5×10^{-4} M. For structures with overhangs, two single-base overhangs were present either at the 3′ or 5′ ends. The 8-mer with a terminal mismatch had a non-complementary base pair at one of the ends of the duplex.

Motif	$T_m - T_m$ (NN^{avg})	T_m (°C)
5-mer	0.6	26.4
6-mer	0.1	42.5
7-mer	−0.1	53.6
8-mer	−0.7	61.2
10-mer	−0.5	72.5
12-mer	−0.9	79.3
6-mer (3′ overhangs)	−1.2	49.8
6-mer (5′ overhangs)	−2.8	43.1
8-mer (3′ overhangs)	−0.7	65.6
8-mer (5′ overhangs)	−3.0	61.9
8-mer (terminal mismatch)	−2.0	56.0

the differences are roughly on the order of the accuracy of the NN-model itself.

To test the sequence-dependent parametrization of the hydrogen-bonding and stacking strength of the interactions, we calculated the melting temperatures for randomly generated ensembles of RNA duplexes, different from the ensemble used for parametrization. A histogram of the differences in the melting temperature predicted by the sequence-dependent version of oxRNA (T_m) and those calculated from the nearest neighbor model ($T_m(NN)$) is shown in Figure 2(a). The main histogram is for duplexes with both Watson-Crick and wobble base pairing, but not containing GU/UG or UG/GU base pair steps. For convenience, the generated ensembles of sequences also do not include any self-complementary sequences because the calculation of their melting temperatures requires a different finite size correction.[73,74] The average difference
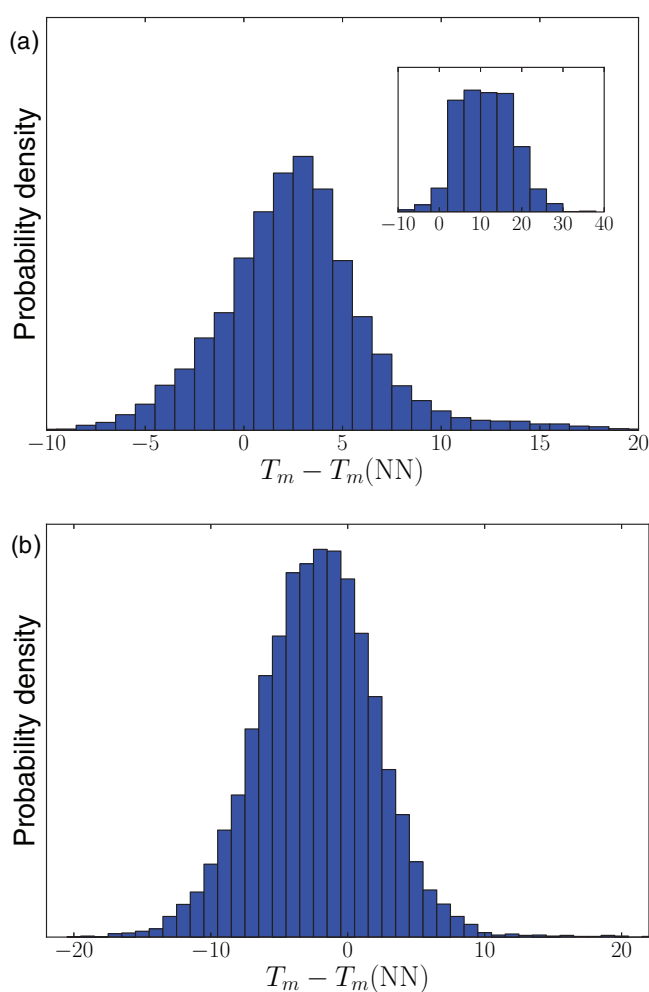

FIG. 2. (a) The histogram of differences between melting temperatures as predicted by the oxRNA model (T_m) and by the NN-model ($T_m(NN)$) for a set of 20 255 randomly generated RNA duplexes of lengths 6, 7, 8, 10, and 12 with Watson-Crick and wobble base pairing. The main plot shows a histogram of values of $T_m - T_m(NN)$ for duplexes that do not include GU/UG or UG/GU base pair steps. The inset shows a histogram of values of $T_m - T_m(NN)$ for 1439 randomly generated sequences that contained at least one GU/UG or UG/GU base pair steps. (b) The histogram of differences between melting temperatures as predicted by the oxRNA model and by the NN-model for a set of 12 000 randomly generated hairpins with stems of lengths 6, 8, and 10 and loops with lengths of 5, 6, 7, 8, 10, and 15, where the stems only contain Watson-Crick base pairs

in melting temperatures is 2.0 °C, with an average absolute deviation of 3.3 °C. The histogram in the inset of Fig. 2(a) is for sequences containing at least one GU/UG or UG/GU base pair step. The average difference in melting temperatures for the ensemble is 9.3 °C and the average absolute deviation is 9.6 °C. We note that in the NN-model, the free-energy contribution of these base pair steps is positive at 37 °C, meaning that they actually destabilize the duplex. However, in the oxRNA model, the cross-stacking, stacking, and hydrogen-bonding interactions are always stabilizing interactions and the interaction strength of two hydrogen-bonded nucleotides does not depend on the identity of their respective neighbors on the strand. Our coarse-grained model hence cannot capture the free-energy contributions of GU/UG and UG/GU base pair steps. One could imagine adding multi-body interactions, but for the sake of computational efficiency and maintaining the consistency of our coarse-graining methodology, we do not do so in this study. Another option might be to introduce a structural perturbation of the helix caused by the GU base pairs.

The histogram in Fig. 2(b) shows the difference between melting temperatures calculated by oxRNA and those predicted by the NN-model for an ensemble of randomly generated hairpins. The average melting temperature difference was −2.8 °C with the average absolute deviation being 4.1 °C.

The transition widths for duplex and hairpin formation were calculated for the averaged model as the difference between the temperatures at which the yield is 0.8 and 0.2, respectively. This quantity is important because the widths of the transition determine the change of the duplex melting temperature with concentration. It can be shown[54] that the derivative of the melting temperature as a function of strand concentration is proportional to the width of the transition. The melting simulations of oxRNA with the average-base parametrization were compared with the width predicted by the averaged NN-model. For the duplexes of lengths 6, 8, 10, and 12 the width was on average underestimated by 0.8 °C, but was overestimated for a 5-mer by 0.5 °C. The width of the transition for the averaged NN-model decreases from 20.5 °C for a 5-mer to 9.2 °C for a 12-mer. For a set of hairpins with stems of length 6, 8, and 10 bp and with loops of lengths 5, 6, 7, 8, 10, and 15, the width of the melting transition was on average underestimated by 1.5 °C. The width of the hairpin transition decreases from about 12 °C for stems of length 6 to approximately 8 °C for stems of lengths 10 in the averaged NN-model. However, the trend of increasing width with decreasing size is always captured by the oxRNA model.

Finally, we note that we could have parametrized the sequence-dependent model only to duplex melting temperatures, as for oxDNA, which would then have led to a larger underestimate of hairpin melting temperatures, presumably because our representation of the strand is too simple to exactly capture the entropy and enthalpy of the loop formation. We hence chose to parametrize to the ensemble of duplexes and hairpins because hairpins are a prominent secondary structure motif in RNA.

Of the other RNA coarse-grained models, the thermodynamics has been most thoroughly studied for the lattice model of Ref. 51, where the model's prediction of melting temperatures was compared with the experimental measurements for ten hairpins and the standard deviation was found to be 4.5 °C. For our model, the standard deviation for the melting temperatures of hairpins considered in the histogram in Fig. 2(b) was 4.3 °C. The model of Ref. 49 was parametrized to reproduce the experimentally measured free-energy stabilization of two hairpins at various salt concentrations but it remains to be seen how the model performs for duplexes or hairpins that were not in the fitting ensemble.

2. Thermodynamics of secondary structure motifs

Given that our aim is to design a model that goes beyond describing hybridization of simple duplexes, it is important to assess how well it is able to reproduce the thermodynamic variation induced by common secondary structure motifs such as bulges or mismatches. To this end, we have studied the melting temperature changes induced by internal mismatches, terminal mismatches and bulges of different lengths.

To assess the effects of bulges, we consider a systems of two strands, one with 10 bases and the other with 10 complementary bases and extra bases that create a bulge motif. We considered bulges of lengths 1, 2, 3, and 6, positioned in the center. For each sequence considered we calculated the melting temperatures by reweighting a set of 6000 states that were sampled from a melting simulation using the average-base parametrization. For each bulge length, we considered 1000 randomly generated sequences with Watson-Crick base pairing in the complementary part.

We further evaluated the melting temperatures for randomly generated 10-mers which contained 1, 2 or 3 consecutive mismatches (and therefore had 9, 8 or 7 complementary Watson-Crick base pairs). The average difference and absolute average deviation for the considered bulges and mismatches are shown in Table II. The melting temperatures of duplexes with bulges are underestimated by a few degrees. However, the model presently significantly overestimates the stability of internal mismatches by the order of 10 °C or more. Even though the mismatching base pairs do not gain stabilization from hydrogen-bonding interactions (which are zero for bases that are not complementary), they still have favorable cross-stacking and stacking interactions, which have their minimum energy configuration in an A-form helical configuration, which the oxRNA model can still form with the mismatches presents. We further note that our model represents each nucleotide by the same rigid body structure, so the mismatching base pairs do not cause any distortion to the duplex structure in our model. Improving the model to more accurately represent the secondary structures with mismatching nucleotides will be the subject of future work.

C. Mechanical properties of the model

1. Persistence length

The persistence length L_p of dsRNA molecule measured in experiments is reported to be between 58 nm to 80 nm,[82–84] corresponding to 206–286 bp (assuming 0.28 nm rise per base pair). The first studies of the persistence length

TABLE II. Melting temperatures for bulge and internal mismatch motifs in a 10-mer. $\langle \Delta T_m \rangle = \langle T_m - T_m(\text{NN}) \rangle$ is the average difference between the melting temperature of the oxRNA model (T_m) and the melting temperature as predicted by the NN-model ($T_m(\text{NN})$). $\langle |\Delta T_m| \rangle = \langle |T_m - T_m(\text{NN})| \rangle$ is the average absolute difference in melting temperatures. $\Delta T_m^{\text{duplex}}(\text{NN})$ and $\Delta T_m^{\text{duplex}}$ are the average differences in melting temperature between the sequences with a secondary structure motif and a duplex with the same sequence but with no bulge or internal mismatch as predicted by the NN-model and oxRNA, respectively. Each of the motifs considered is destabilizing, resulting in a decrease of the melting temperature. The averages were taken over an ensemble of randomly generated sequences (1000 for each motif) that had 10 complementary Watson-Crick base pairs for the bulges, and 9, 8, and 7 complementary base pairs for internal mismatches of size 1, 2, and 3 bases, respectively. The bulges that we consider were of the size 1, 2, 3, and 6 bases. All the melting temperature calculations were calculated for an equal strand concentration of 3.5×10^{-4} M.

Motif	$\langle \Delta T_m \rangle$	$\langle |\Delta T_m| \rangle$	$\Delta T_m^{\text{duplex}}(\text{NN})$	$\Delta T_m^{\text{duplex}}$
Bulge (1 b)	−3.8	4.3	−13.0	−19.5
Bulge (2 b)	−2.1	4.3	−17.5	−22.4
Bulge (3 b)	−4.4	6.0	−19.6	−27.0
Bulge (6 b)	−0.9	5.2	−25.4	−29.1
Internal mis. (1 b)	10.0	10.0	−18.0	−10.9
Internal mis. (2 b)	8.8	9.4	−27.2	−21.2
Internal mis. (3 b)	16.3	16.4	−45.0	−31.5

of dsRNA used electron micrographic, gel-based and hydrodynamic measurements (reviewed in Ref. 83) and reported the persistence length to be between 70 and 100 nm, in salt conditions ranging from 6 mM [Mg^{2+}] and 0.01 M to 0.5 M [Na$^+$]. A more recent single-molecule experimental study[82] in 0.01 M [Na$^+$] and 0.01 M [K$^+$] buffer measured the persistence length by analyzing force-extension curves in magnetic tweezers experiments as well as by analyzing atomic force microscopy images of the RNA duplexes. The two methods yielded consistent values with the measured persistence length estimated as 63.8 and 62.2 nm, respectively, corresponding to 227 and 222 bp. Finally, a recent single molecule force-extension study[84] of dsDNA and dsRNA at salt concentrations ranging from 0.15 M to 0.5 M [Na$^+$] found the dsRNA persistence length to decrease from 67.7 to 57.7 nm with increasing salt concentration, and the extrapolation of measured persistence lengths to higher salt concentrations approaches asymptotically 48 nm (171 bp).

To measure the persistence length in our model, we simulated an 142-bp long double-stranded RNA with the average-base oxRNA model, and measured the correlations in the orientation of a local helical axis along the strand. The local axis vector $\hat{\mathbf{l}}_i$ is fitted through the ith base pair and its nearest neighbors, using the approach described in Sec. S-I of the supplementary material,[64] but considering only the nearest neighbors. We found the results to be robust even when 2 or 3 next nearest neighbors were included in the construction of the local axis. To account for edge effects, a section of ten base pairs at each end of the duplex was ignored when accumulating averages. For an infinitely long, semi-flexible polymer in which the correlations decay exponentially with separation along the strand, the persistence length L_p can be obtained from[85]

$$\langle \hat{\mathbf{l}}_0 \cdot \hat{\mathbf{l}}_n \rangle = \exp\left(-\frac{n\langle r \rangle}{L_p} \right), \qquad (8)$$
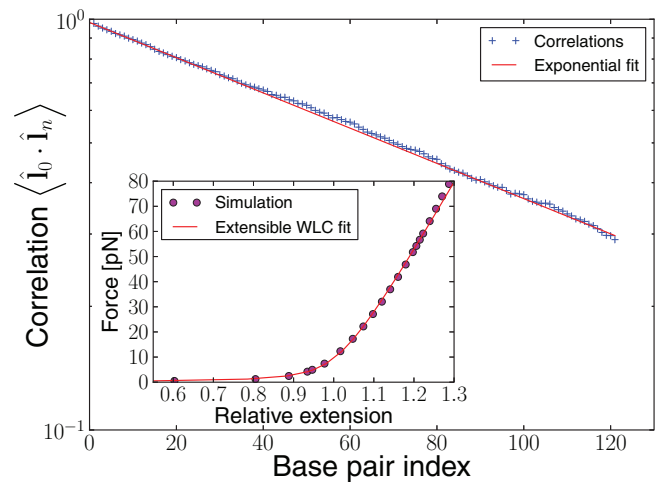

FIG. 3. Semi-logarithmic plot of the correlation function for the direction of the local helix axis along the duplex as defined in Eq. (8) and an exponential fit to the data. The inset shows the force-extension curve of a 81-bp section of a 99-bp duplex. The extension is normalized with respect to the contour length of the 81-bp duplex (with a rise of 2.8 nm per bp) and a fit to the data using the extensible wormlike chain model defined in Eq. (9) is also plotted.

where $\langle r \rangle$ is the rise per base pair. This correlation function is shown in Fig. 3 along with the exponential fit from which we estimated the persistence length of our model to be about 101 base pairs, somewhat lower than the 171 bp persistence length expected at this salt concentration. Our model's persistence length is hence smaller than the experimentally measured values, but still within a factor of 2. OxRNA hence captures the correct order of magnitude for the persistence length and as long as one studies structures whose duplex segments are smaller than the persistence length of the model, it should provide a physically reasonable description.

We note that the persistence length is quite hard to correctly reproduce. We expect this issue to hold for other coarse-grained models as well. To our knowledge, the persistence length has not been measured yet in these models.

2. Force-extension properties

As a further test of the mechanical properties of the model, we measured the extension of a 99-bp RNA duplex as a function of applied force for the average-base model. We used a constant force acting on the center of mass of the first and last nucleotides in one of the two strands in the duplex. We focus on the average extension between the 11th and 91st nucleotide on this strand in order to avoid end effects, such as from fraying of base pairs, in our measurements.

We compare our force-extension data with an extensible worm-like chain model,[86] which provides the following expression for the projection of the end-to-end distance $\mathbf{R}$ along the direction of the force $\hat{\mathbf{z}}$:

$$\langle \mathbf{R} \cdot \hat{\mathbf{z}} \rangle = L_0 \left(1 + \frac{F}{K} - \frac{k_B T}{2 F L_0} (1 + A \coth A) \right), \qquad (9)$$

where

$$A = \sqrt{\frac{F L_0^2}{L_p k_B T}}.$$

Here K is the extension modulus and L_0 is the contour length. This expression comes from an expansion around the fully extended state, and thus it is expected to be valid at forces high enough for the polymer to be almost fully extended.

It was shown experimentally[84] that this extensible worm-like chain model describes the behavior of dsRNA prior to the overstretching transition. In particular, at 0.5 M [Na$^+$], the extensible worm-like chain model fit to the experimentally measured force-extension curve yielded $L_p = 57.7$ nm and $K = 615$ pN.

The force-extension curve for oxRNA is shown in the inset of Fig. 3. We used data from forces between 2.4 pN and 69 pN for our fitting and allowed L_0, K, and L_p to be free parameters. From the fit we obtained $L_0 = 23.4$ nm (84 bp), $K = 296$ pN, and $L_p = 26.0$ nm (93 bp). We note, however, that Eq. (9) is not a robust fit for our model: changing the fitting interval and thus including or excluding points at either high or low forces significantly changes the resulting values of the fitting parameters, even though the residual error in the fit remains small. The persistence length, for instance, can change by more than a factor of two. In the interval we selected for the fit, the L_p obtained approximately corresponds to that obtained from the correlation function in Fig. 3, but given the sensitivity of the fit (which was not observed for the oxDNA force-extension curve[54]), the errors on these extracted parameter values should be taken to be quite large. Another issue to keep in mind is that the inclination angle is also affected by the force. At 10 pN, this is roughly a 1° to 2° change, but close to the point where the chain starts to significantly overstretch (as discussed in Sec. III C 3), the inclination has disappeared, and the bases are almost perpendicular to the axis. It is likely that this deformation is not entirely physical. However, the physical structure of stretched RNA is not experimentally known. In DNA, the structure of the extended state is a very active topic of research.

3. Overstretching

Both DNA and RNA duplexes are known to undergo an overstretching transition at high enough force. Recent experiments[84] for different salt concentrations find 63.6 (2.0) pN for RNA overstretching at 0.15 M [Na$^+$] up to 65.9 (3.3) pN at 0.5 M [Na$^+$]. Following the approach taken in the study of DNA overstretching with the oxDNA model,[62] we used the average-base oxRNA model to run VMMC simulations of a 99-bp RNA duplex with equal and opposite forces applied to the first and last nucleotide of one strand. In our simulations, only native base pairs were allowed to form to aid equilibration, i.e., no misbonds in the duplexes or intrastrand base pairs in the unpeeled strand. The simulations were started from a partially unpeeled state to sample states which have between 65 and 71 bp. The obtained free-energy profiles as a function of the number of base pairs are shown in Fig. 4(a). As the force increases, the slope of the free energy profiles changes from negative (states with more base pairs are favored) to positive (it is favorable for duplexes to unpeel). By estimating the force at which the slope becomes zero, we obtained 86.2 pN as the overstretching force. We note that our model was parametrized for 1 M [Na$^+$], whereas
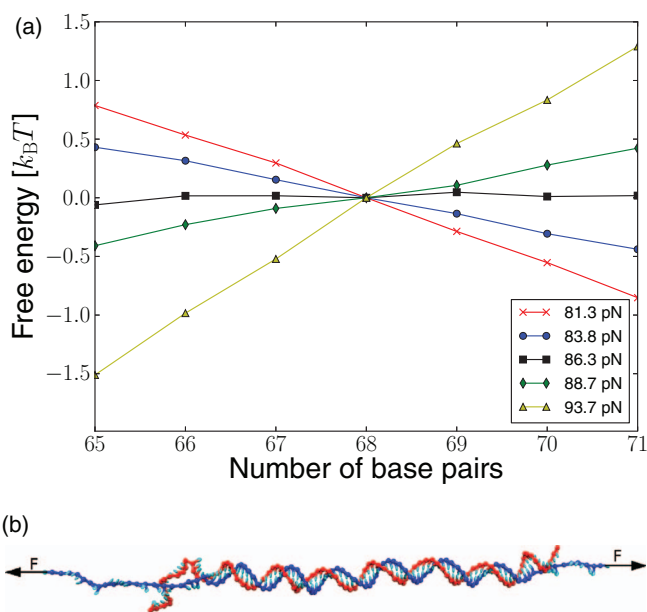

FIG. 4. (a) Free energy as a function of the number of base pairs in the duplex for different forces, where we set the free energy to be 0 for 68 bp for each force considered. At the overstretching force, the slope of the free-energy profile is 0. (b) Snapshot from a VMMC simulation at $F = 86.3$ pN, showing unpeeling from the ends.

the overstretching experiment was done at 0.5 M. Furthermore, by not allowing any formation of secondary structure in the unpeeled strands, we overestimate the overstretching force in the model, because these intramolecular base pairs stabilize the unpeeled state. For the oxDNA model, it was shown that allowing secondary structure decreases the overstretching force by about 3 pN.[62] We would expect the stabilization to be slightly higher for RNA, as RNA base pairs are more stable. Our model hence overestimates the value of the overstretching force by about 16–20 pN. This overestimation is partly due to the higher extensibility of the duplex in oxRNA, which is aided by the loss of inclination in the duplex when higher forces are applied, as we already discussed in Sec. III C 2.

IV. EXAMPLE APPLICATIONS

A. Pseudoknots

Pseudoknots are a common structural motif in RNA. If a strand is represented as a circle and base pair contacts are represented as chords, then its structure is pseudoknotted if the chords cross. Most secondary structure prediction tools do not include pseudoknotted structures in their computations and thus cannot be used to study systems where they are relevant, although some progress has been made in developing efficient algorithms for this task.[87, 88]

Since oxRNA provides an explicitly three-dimensional representation of the RNA strands, it can be used to simulate the folding and thermodynamics of RNA structures that involve pseudoknots. In this section, we use our model to study the well known MMTV pseudoknot.[89] The sequence and the three-dimensional representation of the MMTV
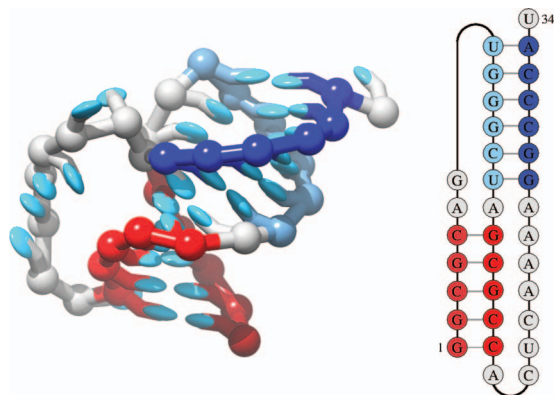
FIG. 5. A snapshot of the MMTV pseudoknot as represented by oxRNA. Stem 1 (shown in blue) has 6 base pairs whereas stem 2 (shown in red) has 5 base pairs. A schematic representation of the secondary structure with the sequence is shown on the right.

pseudoknot by oxRNA are shown in Fig. 5. The MMTV pseudoknot's thermodynamic properties were previously studied in experiment,[89] as well as with another coarse-grained RNA model.[49] Moreover, the MMTV pseudoknot's structure has also been investigated by NMR experiments[90] and two stems were identified in the folded structure: stem 1 with 6 base pairs and stem 2 with 5 base pairs, as schematically shown in Figure 5.

To study the thermodynamics of the system, we ran VMMC simulations of oxRNA for 3.4×10^{11} steps at 75 °C. Umbrella sampling, using the number of base pairs in each of the pseudoknot stems as order parameters, was employed to enhance thermodynamic sampling. We also used histogram reweighting to extrapolate our results to other temperatures. The occupation probabilities of the unfolded state, a single hairpin with stem 1 or stem 2 (denoted as hairpin 1 and hairpin 2), and the pseudoknot are shown in Fig. 6(a). Our simulations also allow us to extract the heat capacity C_V from the expression

$$C_V = \frac{\partial \langle U \rangle}{\partial T}, \tag{10}$$

where we use a cubic interpolation to our simulation data for $\langle U \rangle$ in order to compute the derivative with respect to T. The results are shown in Fig. 6(b).

The experimentally measured C_V at 1 M [Na$^+$] has two peaks, one at 73.5 °C and the other at 95.0 °C.[89] It was hypothesized that the two peaks correspond to the transition from an unstructured strand to a hairpin structure and a second transition from a hairpin structure to the full pseudoknot. Analysis of our yields supports this claim, showing a pseudoknot to hairpin 2 transition at 67.7 °C and transition from hairpin 2 to a single strand with no bonds in stem 1 or stem 2 at 84.8 °C. The higher temperature transition coincides with a peak in the heat capacity, while the lower temperature transition gives rise to a shoulder. While our simulations reproduce qualitatively the behavior of the experimental system, the position of the transitions is not exactly the same as the ones measured experimentally. This is not surprising, as we have shown in Sec. III B that the model generally underestimates the melting temperatures of hairpins.
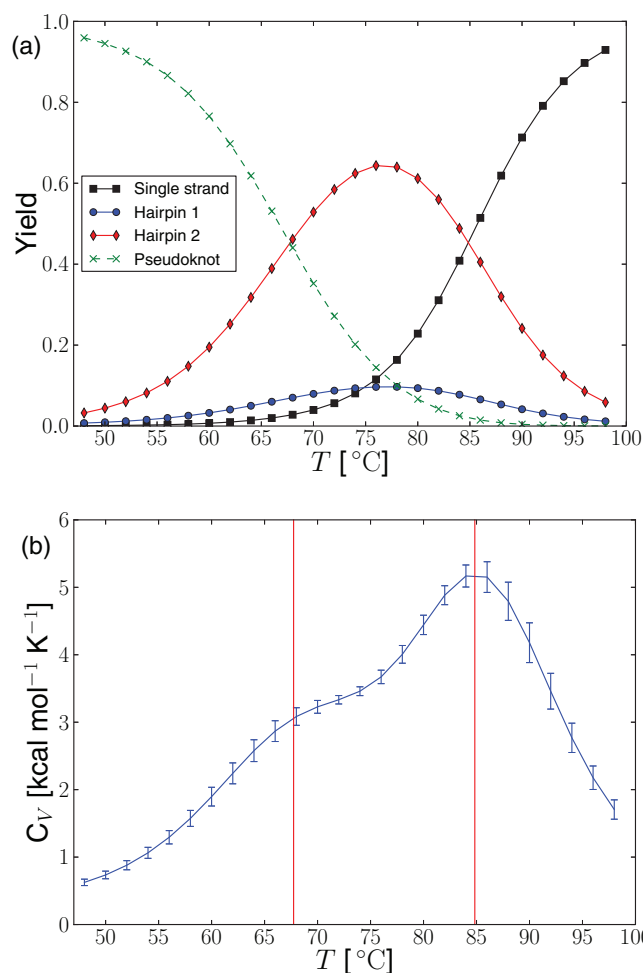

FIG. 6. (a) Equilibrium yields and (b) C_V as a function of temperature for the MMTV pseudoknot. In (a) the pseudoknot and hairpins are defined as having at least 1 native base pair in the relevant stems, whereas the unstructured single-stranded state has no native base pairs. In (b) the error bars are the standard deviations derived from 5 independent simulations. The red vertical lines indicate the temperatures at which we observe equal yields of pseudoknot and hairpin 2 (67.7 °C) and hairpin 2 and the unstructured single strand (84.8 °C).

It is of further interest to analyze the free-energy landscape of the system (Fig. 7). Perhaps unsurprisingly, our results suggest that the minimum free-energy pathway for folding this pseudoknot from a single strand involves first forming one of the stems (forming stem 2 first is more likely as it is more stable and has a higher yield at the considered temperature) and then closing the second stem, instead of simultaneously forming both of them. We have previously seen similar pathways for a DNA pseudoknot.[68]

One subtlety concerns the formation of the GU base pair between the seventh and thirty-fourth nucleotide. The NMR study[90] did not observe the presence of this GU base pair in the pseudoknot structure. However, in our simulations, we find some structures where this base pair forms (thus extending the size of stem 1 from six to seven base pairs), although it has a $5\,k_B T$ free energy penalty at 48 °C with respect to a pseudoknot state which had only six bases in stem 1. Including this additional base-pair within the definition of stem 1 had only a small effect on the calculated yields (the positions
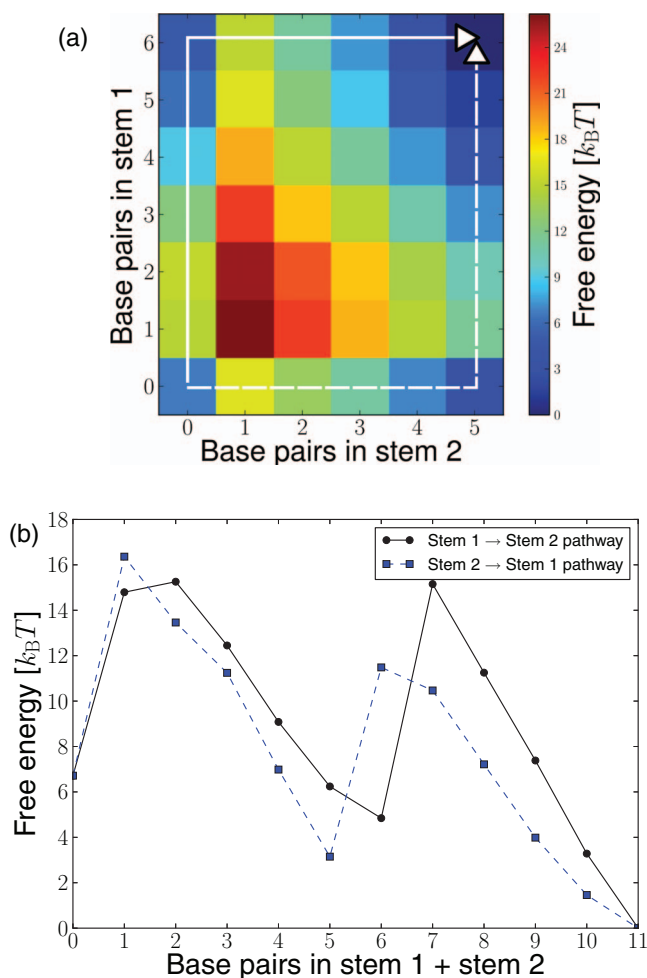
FIG. 7. (a) A free-energy landscape for pseudoknot formation at 48 °C. White lines denote minimum free-energy pathways. (b) Free-energy profiles along the paths indicated in the free-energy landscape of (a). Dashed and solid lines correspond in both pictures. Only native base pairs contribute to the order parameters.

of the equal yields points changed by less than 0.3 °C) and we saw at most 0.5 $k_B T$ free-energy change for the folding pathways. We thus did not include this extra base pair in the definition of stem 1.

The experimental NMR study[90] of the structure of the MMTV pseudoknot found that the two stems of the pseudoknot are bent with respect to each other at about 112°, and the AA mismatch between the sixth and the fourteenth nucleotides to be not stacked. As can be seen in Fig. 5, in oxRNA, this mismatch is typically stacked leading to an angle closer to 160°. We think that this stacking of the stems reflects the overestimation of the stability of mismatches in simpler motifs (see Table II).

In summary, our model is able to describe the thermodynamics of the pseudoknot folding and predict the stabilities of the two stems, supporting the hypothesis that the peak in heat capacity at higher temperature corresponds to the pseudoknot to hairpin 2 transition. The overall secondary structure of the pseudoknot is correct in our model, which also helps to understand the tertiary structure even though we found the angle between the two stems to be larger than the one reported from experiment.

The MMTV pseudoknot thermodynamics was previously studied with the coarse-grained model of Ref. 49 that was able to reproduce the experimentally measured positions of the peaks in the heat capacity curve within about 6 °C precision for 1 M [Na$^+$] and 0.05 M [K^{2+}]. The model of Ref. 49 hence has a better agreement with the experimentally observed transitions than our model, but we note that some of the parameters of the model of Ref. 49 were fitted to the experimentally measured heat capacity curve of the MMTV pseudoknot at 1 M [Na$^+$], whereas our model is fitted to an ensemble of melting data that does not include anything like this MMTV structure.

B. Kissing hairpin complex

A kissing complex is a naturally occurring motif in RNA structures[1] and consists of two hairpins that have complementary loops and can thus bind to each other. An example of such a complex as represented by oxRNA is shown in Fig. 8. The kinetics and thermodynamics of forming a RNA kissing complex with 7 bases in the loops was experimentally studied in Ref. 92 at varying salt concentrations, including 1 M [Na$^+$], the concentration at which our model was parametrized.

To examine the capability of oxRNA to describe kissing complexes, we studied the melting of this kissing complex using both the average-base and the sequence-dependent parametrization of oxRNA and found the transition at a point which is approximately consistent with the observed experimental behavior. The 7-base loops in the two hairpins have
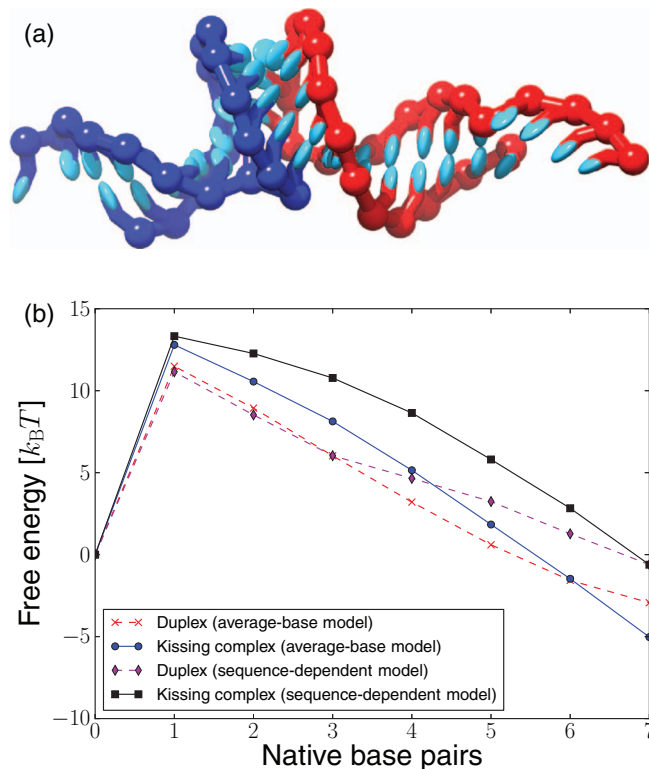

FIG. 8. (a) A typical configuration of a kissing complex between two hairpins that have complementary 7-base loops. (b) Free-energy profiles at 45 °C for forming the kissing complex and a 7-bp duplex with the same sequence as the hairpin loops. Results are shown for the average-base and the sequence-dependent parametrization of oxRNA.

fully complementary sequences (5′-GGAAAUG-3′ and its Watson-Crick complementary sequence). All melting simulations were run in a volume corresponding to an equal strand concentration of 3.5×10^{-4} M.

For the average-base representation we found the melting temperature of the kissing hairpins to be 62.7 °C which compared to 53.6 °C for a 7-bp duplex with the same sequences as the loops. For the sequence-dependent model, we found the melting temperature of the kissing complex to be 44.8 °C, similar to 45.2 °C for this 7-bp duplex. The free-energy profiles for both average-base and sequence-dependent models at 45 °C are shown in Fig. 8.

For most sequences, we find that the kissing hairpin loop is more stable than the separate duplexes with respect to the unbound state. We find that with increasing temperature, the kissing complex loses less stability with respect to the unbound state than a duplex at the same temperature and strand concentration. This trend can be rationalized in terms of the fact that a constrained loop loses less configurational entropy upon binding than an unconstrained single strand does. Furthermore, the kissing complex also gets an extra enthalpic stabilization from a coaxial stacking interaction between the loop and the stem nucleotides. These two effects help explain why, on average, the kissing hairpins are more stable, especially at higher temperatures. However, the kissing hairpins do not satisfy the enthalpic contributions as well as a perfectly formed duplex does. Thus, at low temperature, the duplex can be more stable. Which of these effects dominates depends on the sequence, and if the melting happens before the switch of which motif is more stable, then the duplex will have a higher melting temperature, which we find for a minority of sequences at our strand concentration. For the sequence above, the melting temperatures are very close. Note that the hairpin loops are sufficiently short that kissing complex formation is not inhibited by the topological requirement that the linking number between the loops must remain zero. This contrasts with previous simulations of DNA kissing complexes between hairpins that have 20-base complementary loops.[59]

The thermodynamics of this kissing complexes was studied in Ref. 92 using isothermal titration calorimetry (ITC) at 1 M [Na$^+$]. Up to 35 °C the authors found evidence of a transition to a kissing complex after the injection of the reactants, but did not observe a transition at 45 °C. The authors claim to measure only a 0.6 kcal/mol change in the ΔG for forming the kissing complex between 10 °C and 35 °C while observing a significant increase (19.5 kcal/mol) in ΔH along with a compensating increase in ΔS. Such behavior is not observed in our simulations, where we see a classic quasi-two-state transition in which ΔH and ΔS change slowly with temperature, similarly to a duplex association. If the yields of the kissing complexes in our simulations were extrapolated to the concentrations used in the experiment, we would predict a yield of 35% at 35 °C and 5% at 45 °C. We note that the thermodynamic parameters in Ref. 92 were derived with the assumption that the transition was fully saturated after the injection of the reactants in the ITC experiment, which is incompatible with the experimentally inferred value of ΔG. If the transitions were in fact not fully saturated, then it is

possible that the experiments are consistent with a more typical quasi-two-state transition as observed in our model with a melting temperature approximately consistent with that found by us.

It is also interesting to use oxRNA to probe the structure of this kissing complex because it is a motif that has been used in RNA nanotechnology. Molecular dynamics simulations of the kissing complex using an all-atom representation (AMBER) measured the angle between the helical stems at 300 K and 0.5 M monovalent salt to be approximately 120°.[91] Based on this finding, a hexagonal ring nanostructure that can be assembled from six RNA building blocks was proposed. Each of the proposed building blocks is a RNA strand that in the folded state has a stem and two hairpin loops. The sequences in the loops are designed to bind to the complementary block to allow the assembly of a hexagonal "nanoring" via the kissing complex interaction. This computationally proposed RNA nanoring design was later experimentally realized by self-assembly.[93] The nanoring can be functionalized by including siRNA sequences either in the hairpin stems or by appending siRNA sequences onto the stems. Experiments in blood serum have shown that the nanoring protects the loop regions of the assembly blocks from single-strand RNA endonucleolytic cleavage, making the nanoring a promising tool for *in vivo* nanotechnology applications.[93]

Simulations of oxRNA at 25 °C allowed us to measure the angle between the helical axes of the hairpin stems. We found this angle to fluctuate around an average value of 133.9° with a standard deviation of 14.8°. Hence, an octagon would probably be the most relaxed nanoring for oxRNA, and therefore favored by enthalpy. Smaller rings would be favored by translational entropy.

To illustrate the capabilities of our oxRNA model, we constructed the hexagonal RNA nanoring of Ref. 91 (Fig. 9) by starting a simulation with six folded hairpins blocks and introducing a harmonic potential between the complementary loop regions, which helped the kissing interactions to form. Once the nanoring was completed, the harmonic traps
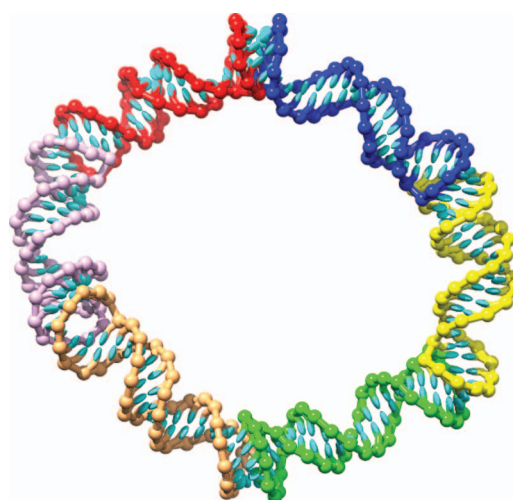

FIG. 9. The hexagonal RNA nanoring of Ref. 91, as represented by oxRNA. The structure is composed of six strands, with a total of 264 nucleotides, connected by kissing loops.

were removed and the assembled structure was relaxed in a molecular dynamics simulation. We found the angle between the stems of the kissing hairpins in the nanoring to fluctuate around a mean of 124.9°. The thermal fluctuations around the mean value had a standard deviation of 14.4°, which is similar to that of the single kissing complex. It would also be possible to use our model to study the mechanical properties of the nanoring, as well as the thermodynamics and kinetics of its self-assembly from the building blocks, but such a study is beyond the scope of this article.

A typical relaxation time for the angle between adjacent kissing complexes or the energy self-correlation function of the assembled nanoring structure corresponds to about a minute of CPU time or less on a standard 2.2 GHz processor. This example shows that structural investigations of systems of the order of hundreds of bases are well within the capabilities of the oxRNA model using a single CPU, and if multiple CPUs are used, or a GPU chip is used,[94] then structural properties and fluctuations around equilibrium can be studied for systems on the order of thousands of base pairs, as can be done for oxDNA.[68]

In Ref. 44, the RNA nanoring considered in this section was studied with coarse-grained models parametrized to reproduce observed angular and distance distributions between groups of atoms in fully atomistic MD simulations of a RNA duplex and the nanoring. The first model proposed in Ref. 44 with a single-bead-per-nucleotide representation was unable to reproduce accurately the nanoring structure as it was found to collapse to unrealistic configurations. In contrast, the second model considered in Ref. 44, with a three-bead-per-nucleotide representation, showed a good agreement with a fully atomistic MD simulation of the nanoring and allows for sampling of the nanoring on the time scale of μs. As our model exhibits a somewhat larger angle between the kissing hairpin stems than the one observed in the MD simulation,[91] the three-bead-per-nucleotide model of Ref. 44 probably provides a more accurate description of the nanoring structure. However, in contrast to our model, the model of Ref. 44 was specifically fitted to the data obtained from the nanoring sampling simulations.

C. Hairpin unzipping

RNA hairpin unzipping has been used in experiment to study the thermodynamics of base pairing and the mechanical properties of RNA, with the kinetics of the process also being of interest.[96–98] Unzipping the same sequence under different salt and temperature conditions can provide systematic data on the physical properties of RNA that, for example, could be used to improve the parametrization of thermodynamic models of RNA.

Here we consider the unzipping of the CD4 hairpin (shown schematically in Fig. 10(a)), which has a 20-bp stem and 4 bases in the loop. It was studied experimentally by pulling at different rates and measuring the unzipping force[96–98] for each trajectory. While it is possible to simulate pulling the hairpins ends at a given rate in the oxRNA model, direct comparisons with experimentally observed unzipping forces are difficult because for a coarse-grained model there
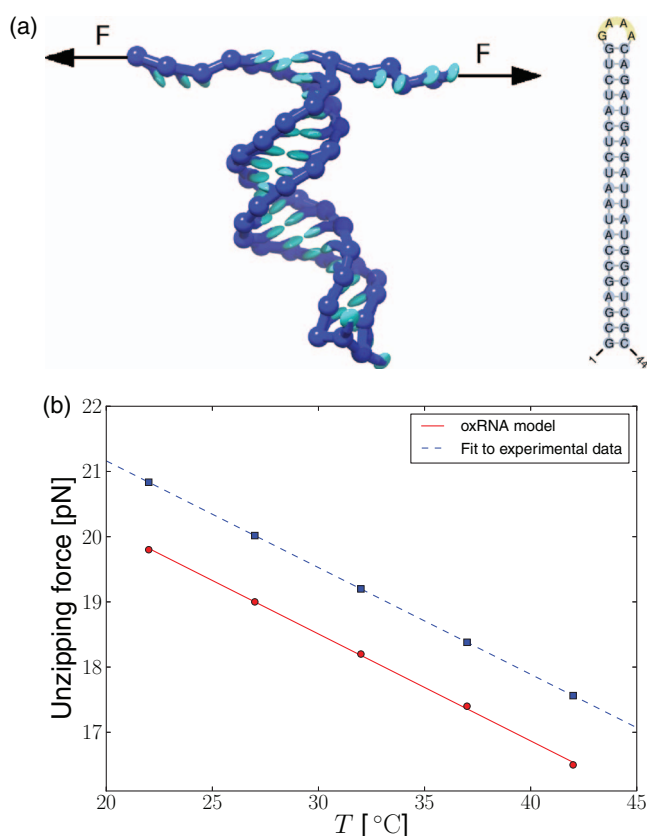

FIG. 10. (a) A snapshot from an oxRNA simulation of unzipping of the CD4 hairpin and a schematic representation of the CD4 hairpin sequence created with the PseudoViewer software.[95] (b) The unzipping forces for the CD4 hairpin as a function of temperature for the oxRNA model (full circles) along with a linear fit to the data. For comparison, the fit to the experimentally measured unzipping force at 1 M [Na$^+$] at temperatures 22, 27, 32, 37, and 42 °C is also shown.

is not a straightforward way to map the simulation time to the experimental one. Furthermore, to obtain an estimate of the average unzipping force for a given pulling rate, one needs to average over multiple trajectories, and generating such trajectories can be quite time consuming, especially for very slow pulling rates.

A more suitable experimental setup for comparison with a coarse-grained simulation comes from Ref. 98, where the authors performed force-clamp experiments at different temperatures and salt concentrations. In the experimental setup, they kept the forces applied to the first and last base of the hairpin constant and measured the folding and unfolding rate of the hairpin, from which they inferred the free-energy difference between the open and closed state. By interpolating results for a range of applied forces, they were able to obtain the unzipping force, which was defined as the force for which the free-energy difference was zero.

To compare to the experimental results, we ran VMMC simulations of the CD4 hairpin with a constant force of 19.7 pN applied to the first and last nucleotide at 23 °C. We then extrapolated the free energies of the open and closed state by the histogram reweighting method to forces ranging from 16.2 to 20.7 pN and to the temperatures at which the experiments were carried out (22, 27, 32, 37, and 42 °C). We only allowed bonds between the native base pairs to avoid

sampling of metastable secondary structures that would slow down our simulations. We considered the hairpin to be closed if at least one of the bonds in the stem was present. For each temperature considered, we performed a linear interpolation of the free-energy difference between closed and open state as a function of force to obtain the unzipping force for which the difference is 0. The unzipping forces we obtain are shown in Fig. 10, along with a fit. We also show for comparison the fit to the experimentally observed unzipping forces[98] at 1 M [Na$^+$], expressed in the form

$$F_{\mathrm{unzip}}(T) = a - cT, \qquad (11)$$

where F_{unzip} is the unzipping force at temperature T. The values obtained in the experiment[98] were $a = 69.1$ pN and $c = -0.164$ pN/K. Fitting Eq. (11) to our simulation data, we obtained the same value for c and 68.2 pN for a. The values of the fitting parameters varied by at most 6% between the fits to unzipping forces obtained from three independent sets of generated states. Thus, oxRNA is able to reproduce the unzipping force with 5% (1 pN) accuracy and fully captures the trend with temperature.

It would be of further interest to study the force-extension properties of a hairpin which contains various secondary structure motifs such as bulges and internal loops or which has regions with variable sequence strengths (such as AU rich and GC rich regions). However, such a study is beyond the scope of this article.

The unzipping of a RNA hairpin was studied with a three-bead-per-nucleotide coarse-grained RNA model in Ref. 50 where the authors produced a free-energy landscape as a function of temperature and force for a P5GA hairpin with 9-bp long stem and 4 bases long loop and also studied the kinetics of unzipping and refolding. The strengths of the interactions in the model of Ref. 50 are set to be proportional to the free-energy values of base pair steps taken from the NN-model, but the thermodynamics of the model was not compared to melting experiments. It would be of interest to see how accurately the unzipping forces of the CD4 hairpin would be reproduced by a recent extension of this model,[49] which has its interactions parametrized to melting experiments. The unzipping of RNA hairpins and larger motifs have also been studied with the SOP model[99] which has a single-bead-per-nucleotide representation. The SOP model was shown to be able to qualitatively reproduce some experimentally observed force extension curves and was used for simulations of unzipping of RNA structures with sizes up to 421 residues.[100] We expect our model to be able to study RNA structures of comparable or even larger sizes. Again, our model was not specifically fitted to reproduce force extension curves or unzipping, but nevertheless appears to reproduce these properties reasonably well.

V. SUMMARY AND CONCLUSIONS

We have presented a new off-lattice coarse-grained model for RNA, oxRNA, which aims to capture basic thermodynamic, structural, and mechanical properties of RNA structures with a minimal representation and pair-wise interaction potentials. OxRNA is specifically developed to allow for efficient simulations of large structures (up to hundreds of nucleotides on a single CPU), and for reactions involving multiple strands of RNA, which are important for applications in RNA nanotechnology. Our coarse-graining strategy is closely linked to our previous successful coarse-graining of DNA with oxDNA.[54] Rather than focusing mainly on reproducing structure, as many other previous models have done, here we tried to systematically compare to a whole suite of different properties, including thermodynamics and mechanics of molecules. Further, we have demonstrated that oxRNA gives a quantitative or semi-quantitative description of a range of systems that are qualitatively distinct from those used for its parametrization. Our model is also currently the only RNA model with reported mechanical properties, which were tested by measuring its persistence length, force-extension curve, and duplex overstretching.

We employed a "top-down" coarse-graining approach, where we aim to reproduce free-energy differences between different states (such as opened and closed state of a hairpin) as measured in experiment. OxRNA represents each nucleotide (i.e., sugar, phosphate, and base) as a single rigid body with multiple interaction sites. The model is capable of representing RNA structures such as hairpins and duplexes and is designed to reproduce the A-form helical RNA duplex. We used a histogram reweighting method[53] to parametrize the model to reproduce the thermodynamics of short duplexes and hairpins. Currently, the oxRNA model captures Watson-Crick and wobble base-pairing interactions as well as various types of stacking interactions. However, it was not designed to capture non-canonical interactions such as Hoogsteen or sugar-edge hydrogen-bonded base pairs, or ribose zippers. Nevertheless, it can reproduce some important tertiary interaction motifs, in particular coaxial stacking of helices, kissing loop interactions, and pseudoknots.

We parametrized oxRNA to reproduce the thermodynamics of the nearest-neighbor model of Ref. 20, which assumes 1M salt concentration. Explicit electrostatic interactions are not included, because they are very short-ranged at high salt and thus can be incorporated into the excluded volume terms in the potential. At lower salt concentration, different longer-ranged forms for the interactions may need to be used to capture the correct physics. We note however that nanotechnology experiments *in vitro* are usually carried out in high salt conditions.

To test our model, we investigated the thermodynamics of short duplexes and hairpins with different sequence content, as well as various other secondary structures such as bulges, internal and terminal mismatches. We found that in comparison with our previous model for DNA, parametrizing RNA thermodynamics is a more challenging task. For example, experimental melting temperatures of a duplex of a given length can differ by as much as 70 °C between weak and strong sequences with Watson-Crick base pairs, as opposed to 50 °C for DNA.[53] So sequence effects are stronger in RNA. Including wobble base pairs presents further challenges, as some base pair steps that include two or more neighboring wobble base pairs have a positive contribution to the free energy of duplex. Although it is not possible to capture this effect with the current representation of our model, adding the

structural effects of wobble base pairs on the double helix may provide means to improve this aspect of the model. Finally, we found that oxRNA overestimates the stability of duplexes with mismatches in internal loops. This could lead to an overestimation of the stability of misfolded structures and complicate the study of the folding of RNA strands that have multiple metastable states with mismatches. Nevertheless, even though oxRNA does not reproduce the exact melting temperatures for structures with internal mismatches and bulges, we do observe, as expected, a decrease in melting temperatures of a duplex with internal mismatches or bulges with respect to the fully complementary strands. The observed changes in melting temperatures are within the orders of magnitude as predicted by the NN-model for the same motifs and capture correctly the direction of the change.

We have tested the mechanical properties of the RNA duplex as represented by the model and found its persistence length to be about half of the reported persistence length of RNA molecules at high salt conditions. The model hence has the correct order of magnitude for the duplex persistence length and provides sufficiently accurate mechanical behavior for most applications, where individual double helical sections are likely to be much shorter than the persistence length.

In Sec. IV, we provided some applications of the model that illustrate its strength and potential utility. In particular, we investigated the thermodynamics of pseudoknot folding and the thermodynamics and structure of a kissing hairpin complex. We also showed that oxRNA can be used to model large nanostructures like a RNA nanoring composed of 264 nucleotides on a single CPU. The computational cost of oxRNA is very similar to oxDNA, where simulations of a DNA origami motif with 12 391 nucleotides have been achieved on GPU.[68] Finally, our model is able to reproduce experimental results for the mechanical unzipping of a hairpin quite closely, to within an accuracy of 1 pN, illustrating oxRNA's potential to study mechanical properties of RNA constructs.

Although we did not describe applications with detailed dynamics (the simulations are typically quite involved and so beyond the scope of this study), we want to emphasize that oxRNA is particularly well-suited for such tasks. For example, oxDNA has been used to study the detailed dynamics of hybridization,[63] toehold-mediated strand displacement,[60] and hairpin formation.[68] Studying similar processes would be very feasible for oxRNA, as would be the self-assembly of RNA nanorings or the kinetics of single-stranded RNA folding.

At this point it is interesting to reflect on some similarities and differences between the coarse-graining of oxDNA and oxRNA. Although oxRNA can clearly reproduce a good number of properties of RNA, quantitative differences with experiment for the melting temperatures of certain motifs are larger than they are for oxDNA. Moreover, it was more difficult to simultaneously reproduce the thermodynamics and the correct persistence length or the force-extension curves. One reason for these differences may be that RNA itself exhibits more complex behavior than DNA, and so is harder to coarse-grain. It is intuitively obvious that the compromises made to increase tractability mean that not all properties of the

underlying system can be simultaneously captured by a more simplified description, a general phenomenon that has been called "representability problems."[101,102] One consequence of representability problems is that in general, fitting too strongly to one set of input data (say structure, as is often done for other RNA models) will often lead to larger errors in other quantities, say thermodynamics. We tried to compromise between different requirements for oxRNA. However, in order to make further progress, more detailed fitting may not be enough. Instead, a more complex and most likely less tractable representation of the full interactions may need to be chosen. For example, for RNA it remains to be seen whether our single nucleotide-level model can be easily extended to generate a better representation of both structure and thermodynamics, or whether, say, a more complex model is needed to achieve the next level of accuracy. Clearly, including electrostatic effects for lower salt-concentrations, or implementing tertiary structure contacts (for example, non-canonical base pairing interactions such as Hoogsteen base pairs or hydrogen bonding between a sugar group and a base) will also need an extension of the current representation.

For RNA models that aim to predict a folded structure, there is a benchmark available by measuring a RMSD for a set of selected structures that are published online.[103] However, there is currently no similar comparison available for thermodynamic properties. If thermodynamics is considered in the design of a coarse-grained model, it is usually reported only for a few systems. Furthermore, as most models are not available for download, it is difficult to compare their computational efficiency. It would be useful in the future to establish a canonical set of thermodynamic data for which coarse-grained models could be compared to one another.

Finally, we note that the simulation code implementing molecular dynamics and Monte Carlo algorithms for our oxRNA model is available for download at dna.physics.ox.ac.uk.

ACKNOWLEDGMENTS

[1] D. Elliott and M. Ladomery, *Molecular Biology of RNA* (Oxford University Press, 2011).
[2] W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
[3] T. Kin *et al.*, Nucleic Acids Res. **35**, D145 (2007).
[4] N. C. Seeman, Annu. Rev. Biochem. **79**, 65 (2010).
[5] J. Bath, S. J. Green, and A. J. Turberfield, Angew. Chem., Int. Ed. **117**, 4432 (2005).
[6] J. Bath, S. J. Green, K. E. Allan, and A. J. Turberfield, Small **5**, 1513 (2009).
[7] P. W. K. Rothemund, Nature **440**, 297 (2006).
[8] L. Adleman, Science **266**, 1021 (1994).
[9] P. Guo, Nat. Nanotechnol. **5**, 833 (2010).
[10] Y. Benenson, Curr. Opin. Biotechnol. **20**, 471 (2009).

[11] K. A. Afonin *et al.*, Nat. Nanotechnol. **5**, 676 (2010).
[12] A. Chworos *et al.*, Science **306**, 2068 (2004).
[13] L. M. Hochrein, M. Schwarzkopf, M. Shahgholi, P. Yin, and N. A. Pierce, J. Am. Chem. Soc. **135**, 17322 (2013).
[14] C. Laing and T. Schlick, J. Phys.: Condens. Matter **22**, 283101 (2010).
[15] C. Laing and T. Schlick, Curr. Opin. Struct. Biol. **21**, 306 (2011).
[16] A. Y. L. Sim, P. Minary, and M. Levitt, Curr. Opin. Struct. Biol. **22**, 273 (2012).
[17] M. J. Serra and D. H. Turner, Methods Enzymol. **259**, 242 (1995).
[18] T. Xia *et al.*, Biochemistry **37**, 14719 (1998).
[19] D. H. Mathews *et al.*, Proc. Natl. Acad. Sci. U.S.A. **101**, 7287 (2004).
[20] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, J. Mol. Biol. **288**, 911 (1999).
[21] A. E. Walter and D. H. Turner, Biochemistry **33**, 12715 (1994).
[22] A. E. Walter *et al.*, Proc. Natl. Acad. Sci. U.S.A. **91**, 9218 (1994).
[23] Z. J. Lu, D. H. Turner, and D. H. Mathews, Nucleic Acids Res. **34**, 4912 (2006).
[24] J. N. Zadeh *et al.*, J. Comput. Chem. **32**, 170 (2011).
[25] I. L. Hofacker *et al.*, Monatsh. Chem. **125**, 167 (1994).
[26] J. S. Reuter and D. H. Mathews, BMC Bioinf. **11**, 129 (2010).
[27] C. Flamm, W. Fontana, I. L. Hofacker, and P. Schuster, RNA **6**, 325 (2000).
[28] A. Xayaphoummine, T. Bucher, and H. Isambert, Nucleic Acids Res. **33**, W605 (2005).
[29] G. Weber, Nucleic Acids Res. **41**, e30 (2013).
[30] J. Šponer, P. Jurečka, and P. Hobza, J. Am. Chem. Soc. **126**, 10142 (2004).
[31] D. Svozil, P. Hobza, and J. Šponer, J. Phys. Chem. B **114**, 1191 (2010).
[32] W. D. Cornell *et al.*, J. Am. Chem. Soc. **117**, 5179 (1995).
[33] B. R. Brooks *et al.*, J. Comput. Chem. **4**, 187 (1983).
[34] G. R. Bowman *et al.*, J. Am. Chem. Soc. **130**, 9676 (2008).
[35] P. Kührová, B. Pavel, R. B. Best, J. Šponer, and M. Otyepka, J. Chem. Theory Comput. **9**, 2115 (2013).
[36] P. Minary, M. E. Tuckerman, and G. J. Martyna, SIAM J. Sci. Comput. **30**, 2055 (2008).
[37] A. Y. Sim, M. Levitt, and P. Minary, Proc. Natl. Acad. Sci. U.S.A. **109**, 2890 (2012).
[38] M. Krepl *et al.*, J. Chem. Theory Comput. **8**, 2506 (2012).
[39] J. Yoo and A. Aksimentiev, Proc. Natl. Acad. Sci. U.S.A. **110**, 20099–20104 (2013).
[40] A. T. Guy, T. J. Piggot, and S. Khalid, Biophys. J. **103**, 1028 (2012).
[41] M. A. Jonikas *et al.*, RNA **15**, 189 (2009).
[42] M. Parisien and F. Major, Nature **452**, 51 (2008).
[43] R. Das, J. Karanicolas, and D. Baker, Nat. Methods **7**, 291 (2010).
[44] M. Paliy, R. Melnik, and B. A. Shapiro, PB **7**, 036001 (2010).
[45] Z. Xia, D. R. Bell, Y. Shi, and P. Ren, J. Phys. Chem. B **117**, 3135 (2013).
[46] S. Pasquali and P. Derreumaux, J. Phys. Chem. B **114**, 11957 (2010).
[47] T. Cragnolini, P. Derreumaux, and S. Pasquali, J. Phys. Chem. B **117**, 8047 (2013).
[48] F. Ding *et al.*, RNA **14**, 1164 (2008).
[49] N. A. Denesyuk and D. Thirumalai, J. Phys. Chem. B **117**, 4901 (2013).
[50] C. Hyeon and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **102**, 6789 (2005).
[51] D. Jost and R. Everaers, J. Chem. Phys. **132**, 095101 (2010).
[52] S. Cao and S.-J. Chen, RNA **11**, 1884 (2005).
[53] P. Šulc *et al.*, J. Chem. Phys. **137**, 135101 (2012).
[54] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Chem. Phys. **134**, 085101 (2011).
[55] T. E. Ouldridge, "Coarse-grained modelling of DNA and DNA nanotechnology," Ph.D. thesis (Oxford University, 2011).
[56] T. E. Ouldridge *et al.*, ACS Nano **7**, 2479 (2013).
[57] P. Šulc, T. E. Ouldridge, F. Romano, J. P. K. Doye, and A. A. Louis, Nat. Comput. (2013); e-print arXiv:1212.4536.
[58] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, Phys. Rev. Lett. **104**, 178101 (2010).
[59] F. Romano, A. Hudson, J. P. K. Doye, T. E. Ouldridge, and A. A. Louis, J. Chem. Phys. **136**, 215102 (2012).
[60] N. Srinivas *et al.*, Nucleic Acids Res. **41**, 10641 (2013).
[61] C. Matek, T. E. Ouldridge, A. Levy, J. P. K. Doye, and A. A. Louis, J. Phys. Chem. B **116**, 11616 (2012).
[62] F. Romano, D. Chakraborty, J. P. K. Doye, T. E. Ouldridge, and A. A. Louis, J. Chem. Phys. **138**, 085101 (2013).
[63] T. E. Ouldridge, P. Šulc, F. Romano, J. P. K. Doye, and A. A. Louis, Nucleic Acids Res. **41**, 8886 (2013).
[64] See supplementary material at http://dx.doi.org/10.1063/1.4881424 for the detailed description of the nucleotide representation and interaction potentials of oxRNA, the description of the fitting of a helical axis to the duplex, and the discussion of the umbrella sampling method.
[65] N. Markham and M. Zuker, "Unafold," *Bioinformatics* (Humana Press, 2008).
[66] I. L. Hofacker, Nucleic Acids Res. **31**, 3429 (2003).
[67] S. Bellaousov, J. S. Reuter, M. G. Seetin, and D. H. Mathews, Nucleic Acids Res. **41**, W471 (2013).
[68] J. P. K. Doye *et al.*, Phys. Chem. Chem. Phys. **15**, 20395 (2013).
[69] S. Whitelam, E. H. Feng, M. F. Hagan, and P. L. Geissler, Soft Matter **5**, 1251 (2009).
[70] T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide* (Springer, 2010).
[71] J. Russo, P. Tartaglia, and F. Sciortino, J. Chem. Phys. **131**, 014504 (2009).
[72] G. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[73] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Phys.: Condens. Matter **22**, 104102 (2010).
[74] T. E. Ouldridge, J. Chem. Phys. **137**, 144105 (2012).
[75] D. H. De Jong *et al.*, J. Comput. Chem. **32**, 1919 (2011).
[76] J. SantaLucia, Jr. and D. Hicks, Annu. Rev. Biophys. Biomol. Struct. **33**, 415 (2004).
[77] D. V. Pyshnyi and E. M. Ivanova, Nucleosides, Nucleotides Nucleic Acids **23**, 1057 (2004).
[78] S. Neidle, *Principles of Nucleic Acid Structure* (Elsevier, 2010).
[79] S. Neidle, *Oxford Handbook of Nucleic Acid Structure* (Oxford University Press, 1999).
[80] P. Kebbekus, D. E. Draper, and P. Hagerman, Biochemistry **34**, 4354 (1995).
[81] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska, Nucleic Acids Res. **37**, 5917 (2009).
[82] J. Abels, F. Moreno-Herrero, T. Van der Heijden, C. Dekker, and N. Dekker, Biophys. J. **88**, 2737 (2005).
[83] P. J. Hagerman, Annu. Rev. Biophys. Biomol. Struct. **26**, 139 (1997).
[84] E. Herrero-Galán *et al.*, J. Am. Chem. Soc. **135**, 122 (2013).
[85] S. F. Edwards and M. Doi, *The Theory of Polymer Dynamics* (Oxford University Press, 1986).
[86] T. Odijk, Macromolecules **28**, 7016 (1995).
[87] E. Rivas and S. R. Eddy, J. Mol. Biol. **285**, 2053 (1999).
[88] M. Bon and H. Orland, Nucleic Acids Res. **39**, e93 (2011).
[89] C. A. Theimer and D. P. Giedroc, RNA **6**, 409 (2000).
[90] L. X. Shen and I. Tinoco, Jr., J. Mol. Biol. **247**, 963 (1995).
[91] Y. G. Yingling and B. A. Shapiro, Nano Lett. **7**, 2328 (2007).
[92] N. Salim *et al.*, Biophys. J. **102**, 1097 (2012).
[93] W. W. Grabow *et al.*, Nano Lett. **11**, 878 (2011).
[94] L. Rovigatti, P. Šulc, I. Z. Reguly, and F. Romano, "A comparison between parallelization approaches in molecular dynamics simulations on GPUs," preprint arXiv:1401.4350 (2014).
[95] Y. Byun and K. Han, Nucleic Acids Res. **34**, W416 (2006).
[96] M. Manosas, D. Collin, and F. Ritort, Phys. Rev. Lett. **96**, 218301 (2006).
[97] C. Bizarro, A. Alemany, and F. Ritort, Nucleic Acids Res. **40**, 6922 (2012).
[98] W. Stephenson *et al.*, Phys. Chem. Chem. Phys. **16**, 906 (2014).
[99] C. Hyeon and D. Thirumulai, Biophys. J. **92**, 731 (2007).
[100] C. Hyeon, R. I. Dima, and D. Thirumalai, Structure **14**, 1633 (2006).
[101] A. A. Louis, J. Phys.: Condens. Matter **14**, 9187 (2002).
[102] A. A. Louis, preprint arXiv:1001.1097 (2010).
[103] J. A. Cruz *et al.*, RNA **18**, 610 (2012).