RESEARCH ARTICLE | MARCH 18 2024

Coarse-grained modeling of DNA–RNA hybrids

Eryk J. Ratajczyk ; Petr Šulc ; Andrew J. Turberfield ; Jonathan P. K. Doye ; Ard A. Louis

J. Chem. Phys. 160, 115101 (2024)
https://doi.org/10.1063/5.0199558
CHORUS

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Coarse-grained modeling of DNA–RNA hybrids

Cite as: J. Chem. Phys. 160, 115101 (2024); doi: 10.1063/5.0199558
Submitted: 17 November 2023 • Accepted: 26 January 2024 •
Published Online: 18 March 2024

Eryk J. Ratajczyk,[1,2] Petr Šulc,[3,4] Andrew J. Turberfield,[1,2] Jonathan P. K. Doye,[5] and Ard A. Louis[6,a)]

AFFILIATIONS

[1] Clarendon Laboratory, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom
[2] Kavli Institute for Nanoscience Discovery, University of Oxford, Dorothy Crowfoot Hodgkin Building, South Parks Road, Oxford OX1 3QU, United Kingdom
[3] School of Molecular Sciences and Center for Molecular Design and Biomimetics, The Biodesign Institute, Arizona State University, 1001 South McAllister Avenue, Tempe, Arizona 85281, USA
[4] School of Natural Sciences, Department of Bioscience, Technical University Munich, 85748 Garching, Germany
[5] Physical and Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QZ, United Kingdom
[6] Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, United Kingdom

[a)] **Author to whom correspondence should be addressed:** ard.louis@physics.ox.ac.uk

ABSTRACT

We introduce oxNA, a new model for the simulation of DNA–RNA hybrids that is based on two previously developed coarse-grained models—oxDNA and oxRNA. The model naturally reproduces the physical properties of hybrid duplexes, including their structure, persistence length, and force-extension characteristics. By parameterizing the DNA–RNA hydrogen bonding interaction, we fit the model's thermodynamic properties to experimental data using both average-sequence and sequence-dependent parameters. To demonstrate the model's applicability, we provide three examples of its use—calculating the free energy profiles of hybrid strand displacement reactions, studying the resolution of a short R-loop, and simulating RNA-scaffolded wireframe origami.

I. INTRODUCTION

DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) are sufficiently similar that they can form stable DNA–RNA hybrids.[1] In a biological context, an important example of such hybrids is the R-loop, which forms when one of the strands in double-helical DNA is displaced by complementary RNA to create a hybrid duplex and an unpaired DNA strand.[2] *In vivo*, short R-loops form during nuclear DNA replication by RNA primers, as well as during transcription when nascent RNA anneals to the DNA template inside an RNA polymerase.[3] The formation of an R-loop is also necessary for the proper functioning of RNA-guided endonucleases in CRISPR-Cas systems, where the guide RNA must fully hybridize with its DNA target for cleavage to take place.[4–6] Much longer R-loops (of the order of 1 kilobase) are formed during the replication of mitochondrial DNA and immunoglobulin class-switch recombination.[3] R-loops play an important role in gene regulation. Errors in their formation and resolution can cause DNA damage, transcription elongation defects, hyper-recombination, and genome instability,[7] and they are also implicated in disease.[8,9] Finally, DNA–RNA hybridization underlies the action of antisense oligonucleotide (ASO) drugs, a therapeutic modality that has shown great promise in, for example, the treatment of neurological disorders.[10–12]

The specificity and predictability of Watson–Crick base-pairing also make DNA and RNA excellent candidate materials for the design of synthetic self-assembled nanostructures, underpinning the growing field of nucleic acid nanotechnology.[13] By simply annealing sets of strands with designed patterns of sequence complementarity, DNA has been used to assemble complex shapes,[14,15] dynamic nanomachines,[16–19] and constructs with potential therapeutic

and diagnostic applications.[20–23] Due to the presence of non-canonical interactions in RNA, its self-assembly is less well characterized. However, the field of RNA nanotechnology is also advancing rapidly, with many examples of functional nanostructures and methods for their assembly.[24–26] The design of nanostructures comprising DNA hybridized to RNA is under-explored, although interest is increasing with exciting potential uses such as the delivery of therapeutic mRNA and artificial ribozyme fabrication.[27–29]

Many different approaches have been developed to tackle the problem of nucleic acid modeling and simulation. Analytical mathematical models such as the worm-like chain (WLC),[30] which treats DNA or RNA as a semi-flexible polymer, can be useful if one is not concerned with details of the structure of the system. Classical molecular dynamics (MD) simulations that consider effective interactions between every atom have yielded useful insights into nucleic acid structure and dynamics, although they can only access microsecond timescales.[31–33] Quantum-chemical calculation is the most fine-grained computational technique used to study nucleic acids,[34] but this is usually limited to very small systems such as dinucleotides.[35] Coarse-grained models, in which groups of atoms are represented as single particles, are a viable intermediate that offers a compromise between speed and detail.[36,37] Multi-scale modeling of DNA nanostructures is reviewed by DeLuca *et al.*[38] While many coarse-grained models of DNA and RNA have been developed,[39–44] modeling of hybrid systems, coarse-grained or otherwise, is relatively sparse and is mostly limited to atomistic simulations.[45–47] Other examples include a mesoscopic model parameterized to reproduce melting temperatures[48] and an abstract model for R-loop formation.[49]

Here, we combine the most up-to-date versions of the models for DNA and RNA developed within the oxDNA framework[50] to enable the simulation of DNA–RNA hybrids. The original average-sequence DNA model[51] has been extended to introduce sequence-dependent thermodynamic properties,[52] improved structural properties, and salt dependence.[53] The same coarse-graining methodology has been used to develop an RNA model.[50,54] A version of the DNA model with sequence-dependent structural and elastic properties is currently under development. The oxDNA family of models has seen tremendous success as tools for the study of nucleic acids and has improved our understanding of DNA and RNA origami[55–61] and strand displacement reactions,[62,63] fundamental nucleic acid biophysics,[64–72] as well as the thermal fluctuation and reconfiguration of flexible DNA nanostructures.[73–75] The introduction of our hybrid model to include DNA–RNA interactions will further expand the range of systems that can be simulated.

II. THE MODEL

Here, we provide a brief overview of the previously developed models for DNA and RNA as well as the introduction of new inter-strand interactions that enable the simulation of hybrids. We then describe in detail how the model was parameterized.

A. oxDNA and oxRNA

In both oxDNA and oxRNA, nucleotides are treated as rigid bodies with interaction sites at the backbone and base. The models take a top-down coarse-graining approach—instead of attempting to exactly replicate the complex intermolecular forces between nucleotides, we use a series of simplified, physically plausible pairwise interactions, which we then parameterize so that our model reproduces the desired properties of the system. The oxDNA/oxRNA interaction potential takes the following form:

$$U = \sum_{bonded} V_{bb} + V_{stck} + V_{exc} + \sum_{nonbonded} V_{HB} + V_{crstck} + V_{cxstck} + V_{DH} + V'_{exc}. \quad (1)$$

The first summation runs over all pairs of particles that are connected through covalent bonds and includes V_{bb}, which enforces backbone connectivity, a stacking potential V_{stck}, and an excluded volume potential V_{exc}. The second summation runs over all remaining pairs of particles—which are not covalently bonded—and includes hydrogen bonding V_{HB}, cross-stacking V_{crstck}, coaxial stacking V_{cxstck}, a Debye–Hückel electrostatic interaction V_{DH}, and excluded volume V'_{exc}. Figure 1 depicts how nucleic acids are represented in oxDNA and indicates the interactions between nucleotides. Coaxial stacking can be thought of as a modified version of the stacking interaction, which is applied across a nick in a backbone. Both the bonded and non-bonded excluded volume terms are implemented as repulsive Lennard-Jones potentials between backbone–backbone, backbone–base, and base–base interaction sites. Details of the exact functional forms of individual interactions can be found in the publications that first introduced the models.[53,54] oxDNA and oxRNA operate within the same framework

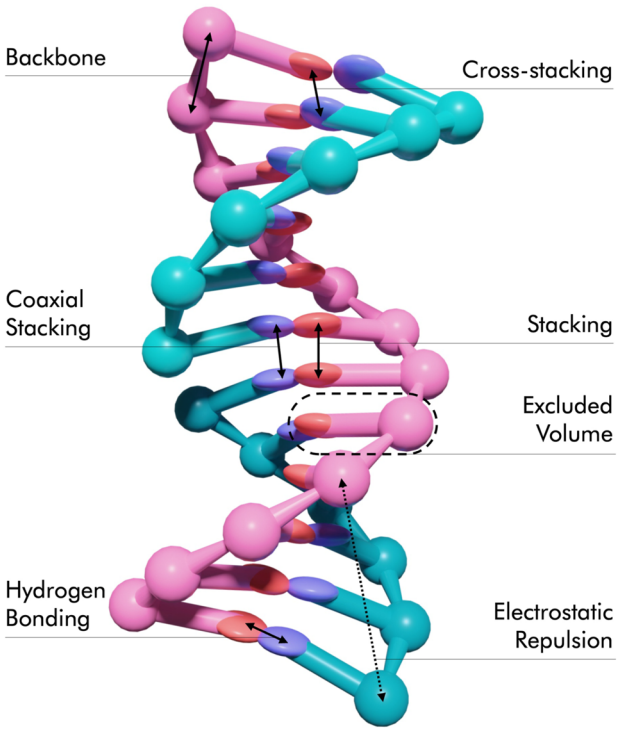

FIG. 1. A nucleic acid duplex, as represented by the coarse-grained model, depicting DNA (blue) hybridized to RNA (pink). Interactions between nucleotides are indicated. Linkages between backbone sites indicate strand directionality, getting thinner in the 5′–3′ direction.

and differ only in the relative positions of interaction sites representing DNA/RNA nucleotides and the parameters that govern the strengths of interactions between them.

B. Incorporating DNA–RNA interactions

The hybrid model was implemented by introducing a new DNA–RNA hybrid potential while using existing potentials to handle DNA–DNA and RNA–RNA interactions. The full interaction potential of our hybrid model for a system containing both DNA and RNA now reads

$$U = U_{DNA} + U_{RNA} + U_{hybrid}. \quad (2)$$

U_{DNA} and U_{RNA} include DNA-only and RNA-only interactions, respectively, and have the same general form as Eq. (1). Interactions between DNA and RNA are represented by U_{hybrid}, which uses non-bonded inter-strand potentials—hydrogen bonding, cross stacking, coaxial stacking, the Debye–Hückel interaction, and excluded volume. Since we do not allow covalent bonds between DNA and RNA nucleotides, U_{hybrid} does not include any of the bonded interactions in Eq. (2). The forms of these new hybrid interactions are the same as those used in the DNA model and, unless otherwise specified, the same parameters were also used.

C. Parameterization

Parameterization of our hybrid model was constrained by the requirement to avoid changes to U_{DNA} and U_{RNA} in order to maintain compatibility with the original oxDNA and oxRNA models; only the hybrid DNA–RNA interactions were modified. In future versions, it would be possible to reparameterize all of U_{hybrid} and potentially obtain an even better fit to experimentally measured properties.

As for oxDNA and oxRNA, we parameterize the hybrid model by fitting it to the predictions of a nearest-neighbor model of thermodynamic properties, which has itself been calibrated to reproduce experimental observations. Nearest-neighbor models for nucleic duplex formation are built by first conducting melting experiments for a range of sequences and using these data to estimate the thermodynamic parameters (ΔH and ΔS) associated with the formation of every possible nucleotide pair in the context of its nearest neighbors (as well as initiation parameters). Using these parameters, one can estimate the melting temperature (T_m) of an entire duplex, which is defined as the temperature at which the single-stranded (ss) and double-stranded (ds) states are equally probable.

Sugimoto *et al.* first estimated nearest-neighbor thermodynamic parameters for DNA–RNA hybrids over two decades ago.[76] A more recent set of improved parameters (now also with sequence-specific initiation parameters)[77] is employed here. The Sugimoto nearest-neighbor model (SNN) predicts melting temperatures to an accuracy of roughly 1 °C; for the purposes of this work, we consider it to be a very good fit to the experiment. Figure 2 highlights the drastic effect that sequence can have on melting temperature in DNA–RNA hybrids, also showing the differences in melting thermodynamics between hybrids, dsDNA and dsRNA. We used the nearest neighbor model of SantaLucia and Hicks[78] to estimate dsDNA melting temperatures and the model of Xia *et al.* for dsRNA.[79] In both cases, we employed an empirical salt correction to T_m derived by SantaLucia.[80] Melting temperatures were calculated using Biopython 1.75.[81] There is a large difference in stability between dA-rU and dT-rA base pairs, which has been attributed to the presence of the C-5 methyl group in thymine and to the location of the 2′-OH group in the purine.[82,83] The stabilities of dG-rC and dC-rG base-pairs also differ, as illustrated by the difference between melting curves for poly-dG-rC and poly-dC-rG. The T_m shown are at a monovalent salt concentration of 1M and a total strand concentration of 3.5×10^{-4}M—values that were also used in melting simulations.

To parameterize our model, we selected hydrogen bonding strength parameters (i.e., potential well depths) for hybrid A–U, A–T, and G–C base-pairs that reproduce the melting temperatures

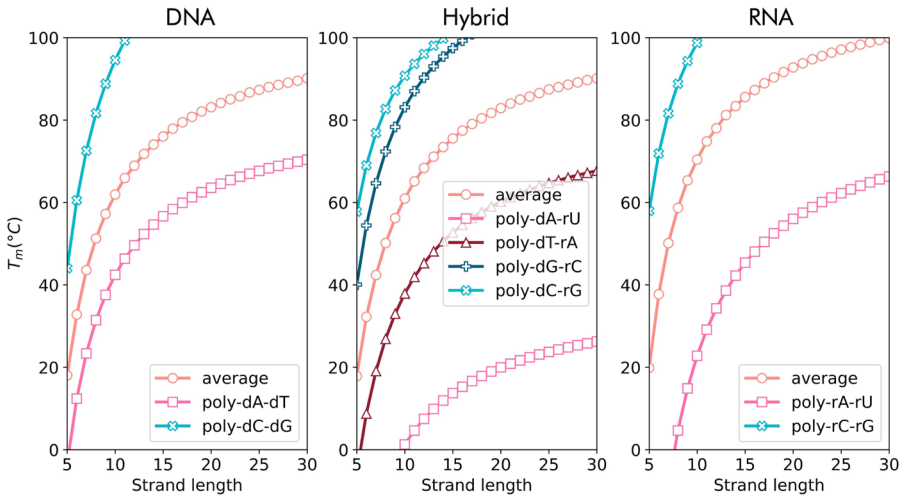

FIG. 2. Duplex melting temperature as a function of strand length for different sequences in DNA, DNA–RNA hybrids, and RNA, as predicted by nearest neighbor models. For the average case, the mean melting temperature of 10 000 random sequences was calculated.

predicted by the Sugimoto model. To estimate the melting temperatures of hybrid duplexes predicted by the model, we simulated duplex dynamics near the melting temperature using the Virtual Move Monte Carlo (VMMC) algorithm[84] and umbrella sampling.[85] Umbrella sampling weights were chosen to ensure that the transition between the single- and double-stranded states was thoroughly sampled. For any given duplex, simulations were run for 10^9 time-steps (three independent simulations for average sequence, one for sequence-dependent versions of the model) at the melting temperature predicted by the Sugimoto model. From these simulations, we obtain the equilibrium populations of single- and double-stranded states, to which we apply a finite-size correction.[86] Using histogram reweighting, one can then extrapolate these populations to the exact T_m, which is defined as the temperature at which the single- and double-stranded states are equally likely—details of the method can be found in the paper introducing oxRNA.[54] We sought a set of parameters that minimized the following cost function:

$$C = \sum_{i \in S} \Delta T_m(i)^2, \tag{3}$$

where $\Delta T_m(i) = T_m^{V\,MMC}(i) - T_m^{SNN}(i)$, $T_m^{V\,MMC}(i)$ and $T_m^{SNN}(i)$ are the melting temperatures predicted by VMMC simulations using our model and the SNN model, respectively, for a given sequence i. The sum runs over a training library, S. For the sequence-dependent parameterization, $S = S_{dep}$, comprising data from 4096 6-mers and 30 000 each of random 8-, 10-, and 12-mers. In the average-sequence parameterization, strand length is the only determinant of T_m, in which case we use $S = S_{avg}$, which only contains four training points: strands of length 6, 8, 10, and 12.

For the average-sequence model, we assume that all hydrogen bonds have the same strength, ε_{HB}. We ran melting simulations for a range of bond strengths for duplexes of lengths 6, 8, 10, and 12—in each case, we found a linear relationship between $T_m^{V\,MMC}(i)$ and ε_{HB}. For each duplex length, we fit a straight line to the data [Fig. 3(a)], enabling an accurate prediction of $T_m^{V\,MMC}(i)$ from ε_{HB}. We chose the value of ε_{HB}, which minimizes C over the average-sequence training library, S_{avg}.

In the sequence-dependent case, we have four possible types of hydrogen bonds, thus four parameters to select: ε_{dArU}, ε_{dTrA}, ε_{dGrC}, and ε_{dCrG}. Note that we distinguish between dG-rC and dC-rG base pairs as well as between dA-rU and dT-rA, as the Sugimoto model suggests that the distribution of bases between the strands affects duplex stability (see Fig. 2). In order to fit sequence-dependent parameters, previous iterations of our coarse-grained models used a histogram reweighting technique to calculate melting temperatures and an annealing algorithm to search the parameter space.[52–54] This approach is necessary when one is fitting >10 parameters, many of which, e.g., stacking strengths, have quite subtle effects on T_m. Since the parameter space to be searched is much smaller, we are able to use a simpler method. We first find an approximate linear mapping between the sequence and the melting temperature predicted by our model. We then use this mapping to find the parameters that best reproduce the melting temperatures predicted by the Sugimoto model.

In order to search the parameter space, we used the following initial minimization procedure: (1) initialize parameters to average-sequence values. (2) For every sequence in S_{dep}, use the current values of ε_{dXrY} to calculate the average bond strength $\bar{\varepsilon}_{HB}(i)$ and, assuming the linear scaling established for the average-sequence model, the corresponding approximation to $T_m^{V\,MMC}(i)$. Compute C. (3) Randomly perturb parameters to generate new parameters, and repeat step 2 for the new parameter set. (4) If new parameters reduce C, accept them; otherwise, repeat step 3. (5) Repeat steps 2–4 until C converges.

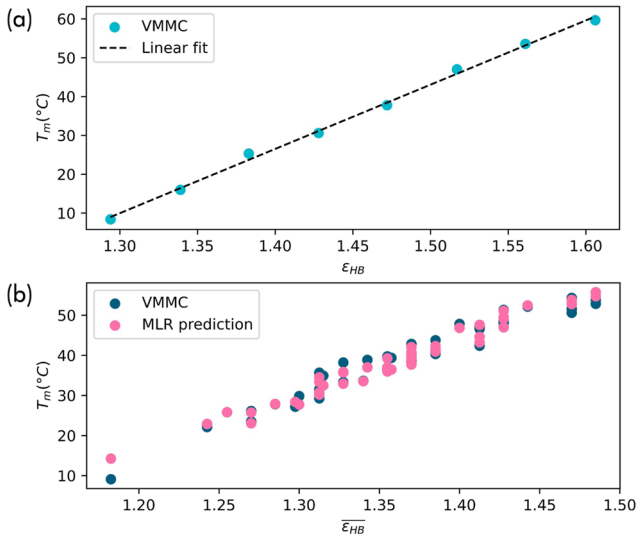

FIG. 3. Dependence of the melting temperature of an 8-mer on hydrogen bonding strength, obtained from VMMC simulations. (a) Melting temperature as a function of hydrogen bonding strength calculated using the average-sequence model alongside a fitted straight line. (b) Melting temperatures of 50 random sequences, calculated by VMMC simulation, plotted against the mean hydrogen bonding strength of the sequence, and the corresponding fits to an MLR model.

In order to further refine the parameters, we needed a better mapping between sequence and $T_m^{V\,MMC}(i)$. With this in mind, we used multiple linear regression (MLR) to predict the result of a VMMC calculation of the melting temperature of a sequence i of length k, such that

$$T_m^{V\,MMC}(i) = \beta_0 + \beta_1 x_1(i) + \cdots + \beta_k x_k(i), \tag{4}$$

where $x_n(i)$ is the hydrogen bonding strength of the nth base-pair within the duplex (read out in a 5′–3′ direction with respect to the DNA strand), and $\beta_0, \beta_1, \ldots, \beta_k$ are fitting parameters obtained from a least-squares minimization. For example, for a hybrid duplex 5′–dATGC–3′/3′–rUACG–5′, we would estimate T_m as $\beta_0 + \beta_1\varepsilon_{dArU} + \beta_2\varepsilon_{dTrA} + \beta_3\varepsilon_{dGrC} + \beta_4\varepsilon_{dCrG}$.

Using our previously obtained estimates of the bonding parameters, we ran melting simulations for 500 random sequences (125 per duplex length) and used these data to fit an MLR model for each length of the duplex [Fig. 3(b)]. We then performed the minimization procedure described earlier—now with the improved VMMC predictions made using the MLR models—to arrive at a refined parameter set. We repeated all of the above (i.e., melting simulations of 500 random sequences, MLR model fitting, followed by minimization) for a final time but found that by this point, the parameters had converged.

Note that initially we also attempted to fit the strength of the cross-stacking interaction for the average-sequence model. We performed preliminary fitting of the hydrogen bonding and cross-stacking (K_{crstck}^{hybrid}) strengths simultaneously and found that the value that minimized C was $K_{crstck}^{hybrid} = 0.938K_{crstck}^{DNA}$, where K_{crstck}^{DNA} is the value used by the DNA model (for reference, $K_{crstck}^{RNA} = 1.262K_{crstck}^{DNA}$). However, we also found that increasing K_{crstck}^{hybrid} while decreasing ε_{HB} accordingly (and vice versa) made little difference to the overall fit. Since the relative strengths of the cross-stacking and hydrogen bonding interactions are not experimentally constrained, different values for K_{crstck}^{hybrid} and ε_{HB} could have been chosen without detriment to the model. We chose to set $K_{crstck}^{hybrid} = 0.938K_{crstck}^{DNA}$ and then selected ε_{HB} using the procedure outlined earlier in this section.

The coaxial stacking and Debye–Hückel interactions could also, in principle, have been reparameterized. However, to the best of our knowledge, no data for DNA–RNA hybrids exist that could be used to fit these interactions. We set the parameters of the Debye–Hückel interaction for hybrids to the values used by the RNA model. The hybrid model uses the same coaxial stacking interaction as the DNA model.

III. PROPERTIES OF THE MODEL

In this section, we report the physical predictions of our model. These include the structure of double-stranded DNA–RNA hybrid duplexes, the melting behavior of both the average-sequence and sequence-dependent versions of our model, and mechanical properties such as persistence length and force-extension characteristics.

A. Structure

The structures of double-stranded DNA and RNA differ significantly—DNA most commonly folds into a B-form helix, whereas RNA takes up an A-form conformation. The A-form helix is characterized by significant slide (displacement of adjacent base pairs along the long axis of the pair) and roll (the angle by which base-pairs open up toward the minor groove), with the result that base pairs are shifted away from the helical axis and inclined to it.[87,88] Figure 4(d) defines the parameters x-displacement and inclination,[89] which are used to characterize the structure of the double helix in this work.

Reports on the exact structure of DNA–RNA hybrids vary. Thanks to studies of polymeric hybrids, it is largely accepted that poly-rA-dT can experience an A- to B-form transition with changes in relative humidity.[90] Hybrids containing poly-dA-rU or poly-dI-rC have been termed *heteromerous*, whereby the DNA and RNA strands possess B- and A-form characteristics, respectively.[91] The detailed structure of oligomeric hybrids can depend on sequence—it is known, for instance, that the purine/pyrimidine content of the DNA strand can change the backbone conformation.[92] A nuclear magnetic resonance (NMR) study by Gyi *et al.*[93] found that the extent of A- or B-form helicity as well as the major/minor groove widths vary with purine/pyrimidine content. They also found that a high-purine DNA strand results in greater conformational diversity as a result of increased sugar flexibility, compared to the case when the RNA strand of the hybrid duplex is high in purine. More recent crystallography studies of oligomeric DNA–RNA hybrids typically characterize them as A-form,[94–96] and estimates of their exact x-displacement and inclination obtained from all-atom simulations suggest a structure in-between those of DNA and RNA.[47]

To determine the structure of a hybrid duplex in our model and compare it to DNA and RNA, we generated 10 000 uncorrelated configurations of each class of a 16 base-pair duplex by performing average-sequence Monte Carlo simulations at 25 °C with a monovalent salt concentration of 0.5M. We then measured the helical parameters of each configuration and calculated their means, as shown in Table I. Note that the values for the RNA model differ from those first reported by Šulc *et al.*[54] since the model used here includes salt-dependent effects that were not included in the original model. Representative structures are shown in Fig. 4. We see that

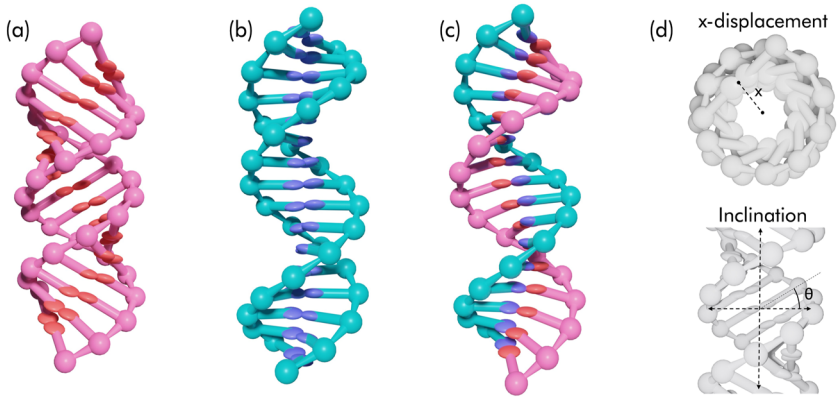

FIG. 4. A comparison of the structures of double-stranded nucleic acids in oxDNA and oxRNA. Shown side-by-side are the structures adopted by a 16-mer of (a) RNA, (b) DNA, and (c) a DNA–RNA hybrid, which naturally adopts a conformation somewhere between that of A-form RNA and B-form DNA. (d) An illustration of the x-displacement and inclination helical parameters, which are used to structurally characterize the helices. The shortest distance, x, between the helical axis and the interaction site where bases meet is defined as x-displacement. The inclination is the angle made between a base and the plane perpendicular to the helical axis, indicated as θ.

TABLE I. Comparison of the inclination, x-displacement, pitch, and rise helical parameters for double-stranded nucleic acids obtained from simulations of our model.

Parameter	DNA	Hybrid	RNA
Inclination (deg)	5.15	8.31	13.8
x-displacement (nm)	0.0536	0.265	0.549
Pitch (bp/turn)	10.6[a]	10.8	11.0
Rise (nm/bp)	0.347[b]	0.343	0.280[c]

[a] As reported by Snodin *et al.*[53]
[b] As reported by Snodin *et al.* at 0.5M salt.[53]
[c] As reported by Šulc *et al.* for the first version of oxRNA.[54]

TABLE II. The hydrogen bonding parameters of the model (in simulation units), compared to analogous parameters for the DNA and RNA models. In the $\varepsilon_{GC/CG}$ row, the hybrid parameters refer to ε_{dGrC} and ε_{dCrG}, respectively.

Parameter	DNA	Hybrid	RNA
ε_{HB}	1.07	1.50	0.87
$\varepsilon_{AU/AT}$	0.89	1.21/1.37	0.82
$\varepsilon_{GC/CG}$	1.23	1.61/1.77	1.06

the values of inclination, x-displacement, and pitch are intermediate with respect to those for DNA and RNA. While our coarse-grained models are not primarily designed to achieve structural accuracy, it is encouraging that the high-level structural features of DNA–RNA hybrids emerge without being explicitly imposed.

In oxRNA, an A-form conformation is imposed on the helix by making the stacking interaction dependent on the angle between the nucleotide orientation vector and the backbone vector connecting neighboring nucleotides, such that the potential energy of the stacking interaction is minimized if the helix adopts an A-form geometry. This angular dependence is not present in the DNA model and, in hybrids, only the RNA strand has this modified stacking interaction. However, the short range of the hydrogen-bonding interaction forces base pairs to lie approximately in the same plane, resulting in a compromise between A- and B-forms. We find that this intermediate helix geometry has an effect on thermodynamic properties, which are discussed in Sec. III B.

B. Thermodynamics

The model parameters selected by the fitting procedure described in Sec. II C and used below are shown in Table II. We note that the hydrogen bonding parameters required to reproduce the correct melting temperatures are substantially larger than in either oxDNA or oxRNA; this point is discussed below.

The fit of the average-sequence model to target melting temperatures is shown in Fig. 5. While, in general, our model reproduces the melting behavior of short hybrid duplexes quite well, there is a noticeable deviation from target temperatures at short strand lengths—for strands of length 6 and 8, the melting temperature is overestimated by around 7.1 and 3.6 °C, respectively. In the average-sequence DNA and RNA models, corresponding deviations are typically no more than 1 °C.

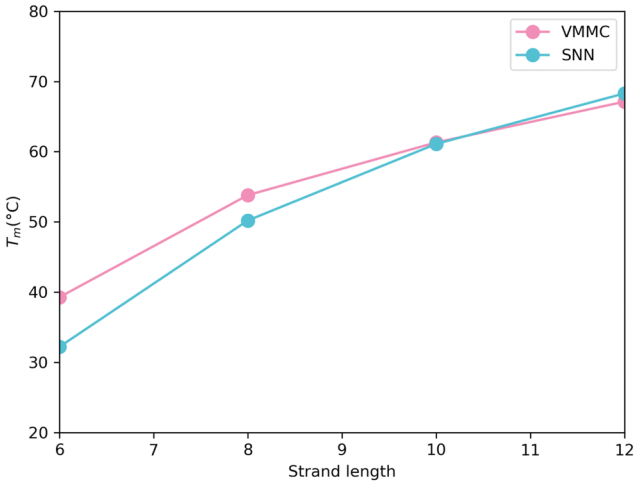

FIG. 5. Melting temperature as a function of duplex length calculated for the average-sequence hybrid model, using VMMC simulations, compared to the target T_m obtained from the Sugimoto nearest neighbor model.

In order to investigate how hybridization between DNA and RNA affects individual interactions, we computed the mean potential energies associated with stacking and hydrogen bonding using a simulation protocol similar to that used in Sec. III A but with the temperature set to 1 °C in order to reduce fluctuations away from the double-stranded ground state. In general, stacking contributes less to the stability of hybrids than dsDNA or dsRNA duplexes. This is because A- and B-form geometries, respectively, were imposed onto the RNA and DNA models through the forms of the interaction potentials: when part of a hybrid duplex, neither the DNA nor RNA is in its preferred conformation, which has a destabilizing effect. This explains why the fitting procedure described in Sec. II C increases the hydrogen bonding strengths to compensate (cf. Table II). We also find that, as strand length increases, both stacking and hydrogen bonding interactions become, on average, less stabilizing. This can be understood as a consequence of stabilizing relaxation of the strained duplex near the ends, which becomes relatively less important as the duplex increases in length. It is also noteworthy that stacking is more disrupted for the RNA strand of a hybrid duplex than for the DNA strand. We propose that this tendency for (RNA) stacking and hydrogen bonding to weaken with increasing strand length is the reason for the melting temperature overestimation in 6- and 8-mers. The model could be further adapted to include a modified stacking potential that can better accommodate hybrids, enabling an even better fit to experimental melting temperatures. This could be implemented by including a double-well angular/radial dependence in the stacking interactions, such that A- and B-form helicities are maintained in dsRNA and dsDNA, respectively, while also allowing a hybrid duplex to inhabit a second potential energy well, mitigating the destabilizing effect in the current version of the model.

In order to test the sequence-dependent version of the model, we ran melting simulations on 1000 random duplexes of lengths 6, 8, 10, and 12 (250 per length). Sequences with predicted melting temperatures below 1 °C (short, U-rich sequences) were discarded.

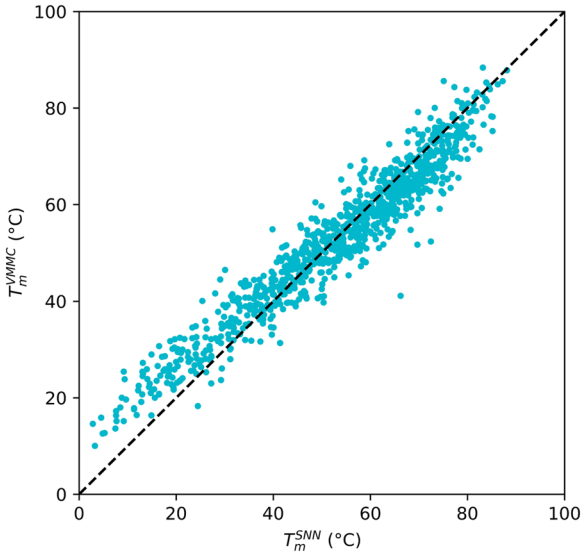

FIG. 6. Performance of the sequence-dependent hybrid model, tested on 1000 random sequences. The plot shows the melting temperature predicted by the model against the value predicted by the Sugimoto nearest neighbor model. The dashed line indicates $y = x$.

The results are shown in Fig. 6. Over this 1000-sequence test set, the model achieves a mean ΔT_m of 0.0926 °C with a standard deviation of 5.36 °C. While we consider this to be a more than satisfactory fit, we are aware of factors that limit our model's performance. The first is its overestimation of the stability of short duplexes, as discussed for the average-sequence model. In Fig. 6, there is a noticeable overestimation of T_m in the <30 °C region, which is almost certainly a manifestation of this effect. As discussed in Sec. III A, sequence can affect backbone conformation. Our model does not factor in these structural changes, which likely worsens the overall sequence-dependent fit.

C. Mechanical properties

The mechanical properties of nucleic acids are biologically important[97] and determine the mechanical behavior of synthetic constructs like DNA origami.[98] For this reason, it is important to check that our model captures the basic mechanics of double-stranded DNA–RNA hybrids. Here, we measure the persistence length and force-extension characteristics of hybrid duplexes within our model and compare the results to available experimental data.

The persistence length L_p of a polymer quantifies its bending stiffness. In a semi-flexible, infinitely long polymer, the persistence length quantifies the correlation between local helix orientations,

$$\langle \mathbf{n}(k) \cdot \mathbf{n}(0) \rangle = \exp\left(\frac{-k\langle r \rangle}{L_p}\right), \tag{5}$$

where $\mathbf{n}(k)$ is the local helical axis vector of the kth base-pair along the duplex and $\langle r \rangle$ is the rise per base-pair.[99] To measure L_p, we performed molecular dynamics (MD) simulations of a 150 base-pair hybrid duplex with the average-sequence model at 22 °C, with the monovalent salt concentration set to 0.5M. We ran ten independent simulations, each for 10^8 time-steps, integrated using Langevin dynamics with a damping constant equal to the time-step. We sampled simulation frames every 10^4 time-steps, giving us a total of 10^5 configurations. For each base-pair, we computed the center of mass and translated it to account for the shift in an A-form helix to give us a point on the helical axis. From these points, we calculate local helical axis vectors, which are used to obtain $\langle \mathbf{n}(k) \cdot \mathbf{n}(0) \rangle$. We discard the five terminal base-pairs to avoid end effects. From the gradient of the line in Fig. 7(a), we obtain an estimate of L_p = 39 nm.

A separate set of simulations was performed to measure the force-extension relationship. We used the same settings as before, except that in this case, we ran 30 replicas, each for 10^7 time-steps. We applied a uniformly increasing, equal, and opposite force of up to 50 pN to terminal nucleotides and sampled the distance between them every 10^3 time-steps to measure the extension, which was averaged over independent simulations. In this case, we fit our data to the extensible worm-like chain model,[100] which predicts that the projected end-to-end distance L of a polymer along the direction of a force with magnitude F is in the limit $F > k_B T/2L_p$,

$$L = L_c\left[1 + \frac{F}{K} - \frac{1}{2}\sqrt{\frac{k_B T}{F L_p}}\right], \tag{6}$$

where K is the stretching modulus and L_c is the relaxed contour length. The results are shown in Fig. 7(b). Fitting to our data gives K = 825 pN, L_c = 51 nm, and L_p = 15 nm. It must be pointed out that the value of K is especially sensitive to the size of the fitting window—for example, fitting up to only 30 pN doubles the estimated stretching modulus (L_p is 25% lower and L_c changes

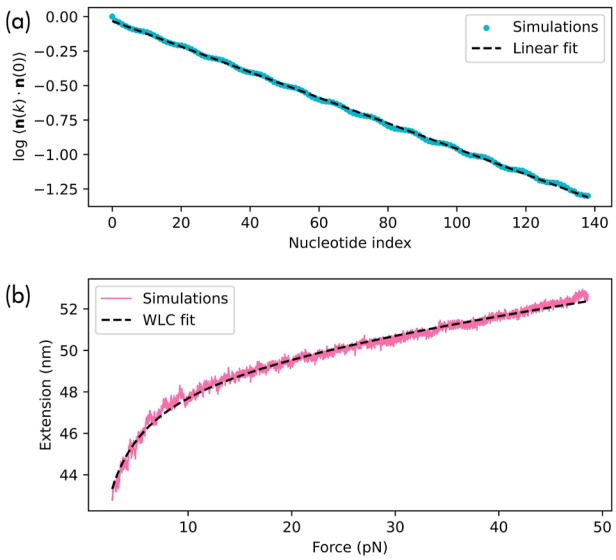

FIG. 7. Measuring the mechanical properties of a 150-mer DNA–RNA hybrid. (a) Natural logarithm of the correlation function $\langle \mathbf{n}(k) \cdot \mathbf{n}(0) \rangle$ against the nucleotide index, k. (b) The force-extension curve obtained from simulations, along with a fit to the extensible worm-like chain.

very little). Note also that a similar issue was observed for oxRNA (but not oxDNA) and was ascribed to a decrease in the inclination angle as the force increased.[54] Consequently, the error on these estimates can be assumed to be relatively large, which should be kept in mind when comparing to experimental values, and the persistence length obtained from the tangent–tangent correlation function should be considered to be more accurate.

Experimental data on the mechanics of hybrid duplexes are scarce, and the number of all-atom simulation studies is also low. Zhang *et al.*[101] performed a series of magnetic tweezer experiments to measure the mechanical properties of a long (>10 kilobase) hybrid duplex at different salt concentrations. They report a stretching modulus of 660 pN, which does not depend strongly on salt concentration. Conversely, salt does have an effect on persistence length, which ranges from 49 to 63 nm at salt concentrations of 0.5 and 1M, respectively. An all-atom simulation study performed at a 1M monovalent salt concentration estimated a stretching modulus of 834 pN.[47] Given that the model is parameterized to reproduce thermodynamic properties, the agreement between calculated and measured elastic properties is satisfactory. We note that for low applied forces (<35 pN or so), the persistence length is more significant than the stretching modulus in determining the mechanical behavior of the duplex.

To put this into perspective, the persistence length L_p of dsDNA at moderate to high salt concentration is in the range 45–50 nm, and the stretching modulus K is around 1050–1250 pN at high salt.[51] The first version of the oxDNA model achieves L_p = 43.8 nm and K = 2120 pN. For dsRNA, experimental estimates of L_p are in the range 58–80 nm and K = 615 pN, while for the oxRNA model, L_p = 28.3 nm and K = 296 pN.

IV. APPLICATIONS OF THE MODEL

We provide examples of the application of the model to three hybrid systems—toehold-mediated strand displacement (TMSD), a short R-loop, and RNA-scaffolded wireframe origami, all of which are technologically and/or biologically important.

A. Toehold-mediated strand displacement

Toehold-mediated strand displacement (TMSD) is a process in which one of the strands within a nucleic acid duplex is exchanged for another. The displacement of the incumbent strand is initiated by the binding of the invader to a short single-stranded toehold region on the complementary strand[16,102] [Fig. 8(a)]. TMSD has many applications in nanotechnology, including in the construction of synthetic molecular circuits.[103] DNA–RNA hybrid TMSD is of particular interest by virtue of its relevance to *in vivo* applications.[104] Strand displacement has also been argued to play an important role in various naturally occurring RNA systems.[105]

We note that oxDNA has been remarkably successful in reproducing experimental observations related to TMSD, having been used, for example, to study mismatches as a tool for modulating strand displacement kinetics.[106,107] RNA strand displacement has likewise been simulated using the oxRNA model.[62]

Here, we use our newly developed model to study strand displacement systems involving DNA–RNA hybrids. As in the melting simulations, we use a combination of VMMC and umbrella sampling to explore the state space efficiently. From the simulations, we obtain unbiased estimates of equilibrium populations of states parameterized by the number of substrate-invader hydrogen bonds. In simulations of toehold-mediated strand displacement, we assigned a weight of zero to states with no hydrogen bonds between substrate and invader or substrate and incumbent hydrogen bonds to prevent dissociation. Umbrella sampling weights were chosen (by trial and error) so that all states have approximately equal occupancy (within an order of magnitude) in the biased ensemble. Assuming that the state space has been adequately sampled, the free energy difference ΔG between states A and B can be written as

$$G(A) - G(B) = -k_B T \ln\left[\frac{p(A)}{p(B)}\right], \tag{7}$$

where $p(A)$ and $p(B)$ are the probabilities of being in states A and B, respectively. We can similarly compute free energy profiles for systems with multiple states. For every system studied, we ran ten

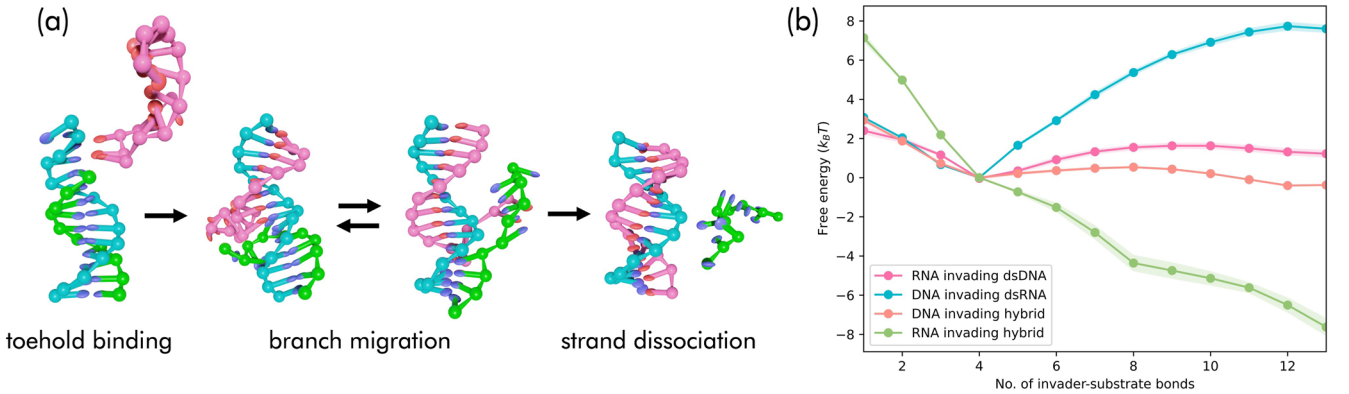

FIG. 8. Simulation of toehold-mediated strand displacement using the average-sequence model. (a) Snapshots of key steps for one of the strand displacement reactions: An RNA strand (pink) invades a DNA duplex (blue and green) by binding initially to a single-stranded toehold. (b) Free energy profiles of the different systems simulated, showing the free energy (set to zero for a fully occupied toehold) against the number of hydrogen bonds between the substrate and invader strands. Error bars indicate the standard error of the mean.

independent simulations for 10^9 time-steps each at 37 °C and a 0.5M monovalent salt concentration using the average-sequence model. We simulated four systems—an RNA strand invading dsDNA, a DNA strand invading dsRNA, a DNA strand invading a hybrid duplex to displace an RNA incumbent from a DNA substrate, and finally an RNA strand invading a hybrid duplex to displace an DNA incumbent. In each case, the toehold region was four nucleotides long, with a ten-nucleotide branch migration domain. Results are shown in Fig. 8(b).

A common feature of all of the free energy profiles is the initial downhill trajectory in the range of 1–4 invader-substrate hydrogen bonds. This is associated with toehold binding, which is always favorable as there is no competition between strands. Generally, there is an entropic barrier associated with the formation of a branch junction during strand displacement, which is seen as an activation barrier in the branch migration region (for RNA invading dsDNA and DNA invading hybrids). In the case of DNA invading dsRNA, the landscape is steeply uphill, as on average, dsRNA is substantially more thermally stable than a DNA–RNA hybrid. Conversely, when RNA invades a hybrid, this results in the formation of dsRNA, which is much more thermally stable than a hybrid duplex, resulting in a downhill landscape. This can be understood in terms of the difference in average melting temperature between dsDNA and dsRNA—around 60 and 71 °C for a ten base-pair duplex, respectively. The difference between the free energy landscapes for RNA invading dsDNA and DNA invading a hybrid is more subtle because hybrids and dsDNA are quite close in melting temperature (around 61 °C for a ten base-pair hybrid duplex). It is likely that this relative difference is smaller than the typical effects of varying base sequences.

The simulations performed here only scratch the surface of what can be studied with the model—future work will investigate the effect of sequence on TMSD free energies and kinetics. Preliminary simulations with the model suggest that free energy landscapes, as well as reaction kinetics, are strongly sequence-dependent. We are also looking into how the secondary structure in the RNA strand impacts the reaction. Given the success of previous oxDNA models in studying TMSD, we are confident that our DNA–RNA hybrid model will provide useful insights.

B. R-loop resolution

An R-loop is a three-stranded nucleic acid structure consisting of double-stranded DNA that is partially hybridized with complementary RNA. As discussed in Sec. I, this is possibly the most important naturally occurring DNA–RNA hybrid system.

We use our coarse-grained model to simulate the resolution of an R-loop. While this system appears to be similar to the TMSD studied in Sec. IV A, as both involve DNA–RNA strand displacement, we observe behavior that is quite different. The simulation protocol used closely resembles our TMSD simulations. We study a single R-loop consisting of 55 base-pair double-stranded DNA that is hybridized to a 25-nucleotide RNA strand at its center [Fig. 9(a), top]. As before, in order to prevent strand dissociation, we restrict the system to states with at least one DNA–DNA and one RNA–DNA hydrogen bond and use average-sequence parameters. In this case, we ran separate simulations for two overlapping windows of the order parameter space—one restricted to 1–13 RNA–DNA hydrogen bonds and another to 13–25 bonds. We performed ten independent VMMC simulations per window, each for 3×10^8 time-steps. Temperature and monovalent salt concentration were the same as for our TMSD simulations.

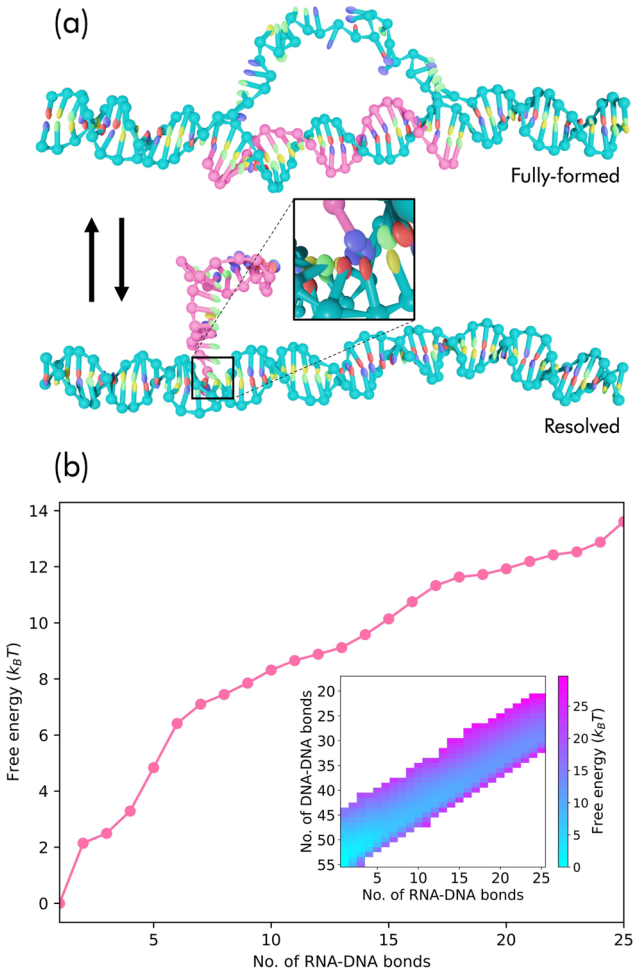

FIG. 9. Studying the resolution of a short R-loop. (a) A fully formed 25-nucleotide R-loop consisting of double-stranded DNA (blue) and a single strand of RNA (pink). Through the process of strand displacement, the system can resolve the R-loop by forcing out the RNA strand. In our simulations, this transition is sampled many times in both directions. (b) Free energy of the system as a function of the number of RNA–DNA hydrogen bonds and (inset) the number of both DNA–DNA and RNA–DNA bonds (states with fewer than a total of ~45 base pairs are not sampled).

Computed free energy profiles are shown in Fig. 9(b). There is a barrier of around $2k_BT$ associated with the transition from 1 to 2 RNA–DNA bonds. The zoomed-in snapshot of the resolved state in Fig. 9(a) suggests an explanation. In the resolved state, the DNA double helix tends to be fully closed, with the RNA strand forming a weak hydrogen bond with one of the DNA strands. As a result, in order to make the transition from one to two RNA–DNA bonds, two DNA–DNA bonds must be broken, which is energeti-

cally costly. This is, in part, an artifact of restricting the simulation to bound states. Without this restriction, the RNA strand would have dissociated completely in the resolved state.

In general, we observe that the formation of the DNA–RNA hybrid in this particular system is significantly less favorable than in the analogous TMSD reaction of RNA invading dsDNA, depicted in Fig. 8(b). Several factors contribute to the difference between the two energy landscapes. In a fully formed R-loop, displacement of the RNA strand can take place from either end; the DNA loop is tethered at both sides, increasing its proximity to the hybrid and making displacement more likely. Resolving an R-loop is clearly entropically favorable, as it entails the exchange of a single strand tethered at both ends for one tethered at only one end in our simulations or fully displaced in practice, thus having much greater conformational freedom.

We also observe an oscillatory component to the free energy, which has minima at R-loop sizes of around 13 and 23 RNA–DNA bonds. When the DNA–RNA hybrid helix is of a size roughly commensurate with its pitch (around 11 base pairs), the ends of the displaced DNA loop are on the same side of the duplex, which entails higher conformational freedom. Conversely, at half a turn away, e.g., around 18, the ends are at opposite sides of the duplex, reducing conformational freedom and leading to a slight additional increase in free energy cost.

The inset in Fig. 9(b) depicts a 2D free-energy landscape that provides additional information about the system. The presence of the R-loop destabilizes the DNA double helix beyond the region of the DNA–RNA hybrid, with states that are not fully hybridized being readily accessible. This is clear from the fact that, at any given number of RNA–DNA bonds, states with numbers of DNA–DNA bonds below what would be expected for a fully hybridized system (55 bonds in total) are sampled.

The stability of an R-loop depends on its length and sequence.[108] An obvious future application of our model would be a comprehensive study of the effects of these factors. The kinetics of R-loop resolution could also be studied using specialized sampling techniques.

C. RNA-scaffolded wireframe origami

Nucleic acid origami is one of the most common techniques used for assembling single-stranded DNA/RNA building blocks into a target structure. Origami nanostructures consist of a scaffold, which is a long strand running through the entire assembly, and shorter staple strands that hybridize into two or more scaffold domains to control their spatial arrangement. Domains of the scaffold strand, which are widely separated in their primary sequence, can be held in close spatial proximity in the final structure. This technique has been applied primarily to DNA, although interest in the design of DNA–RNA hybrid nanostructures is increasing.

We have used our model to simulate three hybrid wireframe origami nanostructures from Parsons *et al.*,[29] which consist of an RNA scaffold and DNA staples. The structures were designed assuming a double helix with a pitch of 11 base-pairs per turn, which is roughly reproduced by our model. We performed MD simulations at 4 °C and a monovalent salt concentration of 0.3M to match the experiments. Each structure was simulated for 10^7 time-steps, and the positions of particles were sampled every 10^4 time-steps for analysis. We simulated three nanostructures—a tetrahedron, an octahedron, and a pentagonal bipyramid—each having edges 66 base-pairs long. For each, we calculated the mean structure and per-nucleotide RMSF (root-mean-square fluctuation). From these mean structures, we reconstructed all-atom models of the nanostructures using the oxDNA-to-PDB converter on TacoxDNA[71] (by superim-

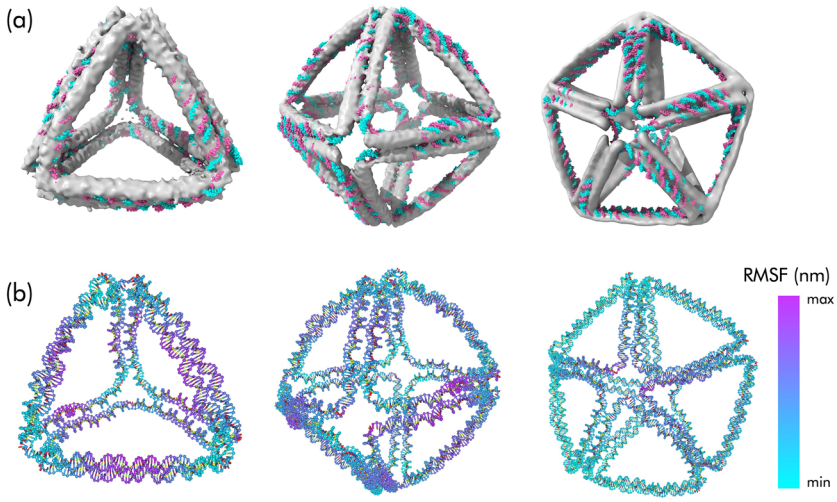

FIG. 10. Mean structures of RNA-scaffolded origami simulated using the model. (a) Atomic models of the tetrahedron (left), octahedron (middle), and pentagonal bipyramid (right), each consisting of an RNA scaffold strand (pink) and DNA staples (blue). Experimentally obtained cryo-EM densities (gray) have been superimposed onto each structure. (b) Structures with coloring to indicate the per-nucleotide RMSF. The structures have different fluctuation ranges: 1.15–1.72, 1.38–2.13, and 1.47–3.87 nm, respectively.

posing atomic coordinates onto individual nucleotides) and then aligned them with cryo-EM densities, obtained by Parsons *et al.* and retrieved from EMDB,[109] using ChimeraX.[110]

Figure 10 compares our results to the experimental data. Our model captures the measured structures reasonably well, with no systematic strain build-up. For the tetrahedron and octahedron, it is immediately clear that structural fluctuations are concentrated at edge centers.

V. CONCLUSION

We have introduced a new coarse-grained model based on existing oxDNA and oxRNA models that enables the simulation of DNA–RNA hybrids. As with previous models, we parameterized the hydrogen bonding interaction to reproduce the melting temperatures of short duplexes. Quantitative agreement with the experimentally calibrated nearest-neighbor model of the thermodynamics of hybrid duplexes is nearly as close as that achieved for DNA and RNA duplexes using oxDNA and oxRNA. The persistence length and stretching modulus derived from simulations of longer duplexes are consistent with experimental values, although some uncertainty about their values remains. The conformation of DNA–RNA hybrid duplexes is a compromise between the structures preferred by DNA and RNA alone. As a result, the stabilization of the duplex by stacking interactions is reduced, necessitating an increase in hydrogen bonding strength to produce the desired melting temperatures. One consequence of this choice is that the model overestimates the stability of short double-stranded helices—something that users of the model should keep in mind. Nevertheless, the overall performance of our DNA–RNA hybrid model for the systems we studied gives us confidence that it will be able to capture the sequence-dependent kinetics/thermodynamics of more complex biophysical processes. A future version of the model will include a modified stacking potential that can accommodate the preferred conformations of dsDNA, dsRNA, and DNA–RNA hybrids.

We have demonstrated the versatility and applicability of our model by performing simulations for three different systems. Our study of toehold-mediated strand displacement using the average-sequence model suggests that the relative stabilities of DNA–DNA, RNA–RNA, and DNA–RNA duplexes play a key role in determining the free energy landscapes of hybrid displacement reactions. Our simulations show that the biophysics of R-loop resolution includes geometric effects related to the commensurability of the R-loop length and the pitch of the double helix. Finally, we have shown that our model can help validate DNA–RNA hybrid origami designs.

Future work will focus on DNA–RNA hybrid systems at time and length scales that are inaccessible to all-atom simulations, including the sequence-dependent kinetics of strand displacement reactions and the effects of RNA secondary structure motifs.

SUPPLEMENTARY MATERIAL

In the supplementary material, we provide values of the β parameters from Eq. (4), as well as the DNA sequence used in Sec. IV B.

ACKNOWLEDGMENTS

The authors acknowledge Thomas Ouldridge and Jonathan Bath for useful discussions, Lorenzo Rovigatti and Erik Poppleton for their help with code development, and Erik Winfree for suggesting the name oxNA for the model. E.J.R. acknowledges the financial support provided by the Clarendon Fund, Somerville College (Oxford), and the Engineering and Physical Sciences Research Council (Grant No. EP/W524311/1). We also acknowledge the Advanced Research Computing service at the University of Oxford for computer time. P.Š. acknowledges support from the National Science Foundation under Grant No. CCF 2211794.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Eryk J. Ratajczyk: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Petr Šulc**: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Project administration (equal); Supervision (equal); Writing – review & editing (equal). **Andrew J. Turberfield**: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Project administration (equal); Supervision (equal); Writing – review & editing (equal). **Jonathan P. K. Doye**: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Project administration (equal); Supervision (equal); Writing – review & editing (equal). **Ard A. Louis**: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Project administration (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The code implementing the model, along with the supporting documentation, can be found at https://lorenzo-rovigatti.github.io/oxDNA/. A new topology file format supporting DNA–RNA hybrids has been implemented in the official oxDNA code, and the accompanying suite of analysis tools has likewise been extended to enable the analysis of systems containing both DNA and RNA. The online visualization tool oxView.org[111] has been extended to also support viewing of DNA–RNA hybrids. The simulations performed here were run on single CPUs, although a GPU version of the model is a likely future development. The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] G. Milman, R. Langridge, and M. J. Chamberlin, "The structure of a DNA-RNA hybrid," Proc. Natl. Acad. Sci. U. S. A. **57**, 1804–1810 (1967).

[2] E. Petermann, L. Lan, and L. Zou, "Sources, resolution and physiological relevance of R-loops and RNA–DNA hybrids," Nat. Rev. Mol. Cell Biol. **23**, 521–540 (2022).

[3]A. Aguilera and B. Gómez-González, "DNA–RNA hybrids: The risks of DNA breakage during transcription," Nat. Struct. Mol. Biol. **24**, 439–443 (2017).
[4]B. Zhang, D. Luo, Y. Li, V. Perčulija, J. Chen, J. Lin, Y. Ye, and S. Ouyang, "Mechanistic insights into the R-loop formation and cleavage in CRISPR-Cas12i1," Nat. Commun. **12**, 3476 (2021).
[5]M. Pacesa, L. Loeff, I. Querques, L. M. Muckenfuss, M. Sawicka, and M. Jinek, "R-loop formation and conformational activation mechanisms of Cas9," Nature **609**, 191–196 (2022).
[6]F. Jiang and J. A. Doudna, "CRISPR–Cas9 structures and mechanisms," Annu. Rev. Biophys. **46**, 505–529 (2017).
[7]C. Niehrs and B. Luke, "Regulatory R-loops as facilitators of gene expression and genome stability," Nat. Rev. Mol. Cell Biol. **21**, 167–178 (2020).
[8]C. Rinaldi, P. Pizzul, M. P. Longhese, and D. Bonetti, "Sensing R-loop-associated DNA damage to safeguard genome stability," Front. Cell Dev. Biol. **8**, 618157 (2021).
[9]A. Brambati, L. Zardoni, E. Nardini, A. Pellicioli, and G. Liberi, "The dark side of RNA:DNA hybrids," Mutat. Res., Rev. Mutat. Res. **784**, 108300 (2020).
[10]D. Di Fusco, V. Dinallo, I. Marafini, M. M. Figliuzzi, B. Romano, and G. Monteleone, "Antisense oligonucleotide: Basic concepts and therapeutic application in inflammatory bowel disease," Front. Pharmacol **10**, 305 (2019).
[11]J. Lee and T. Yokota, "Antisense therapy in neurology," J. Pers. Med. **3**, 144–176 (2013).
[12]C. Rinaldi and M. J. A. Wood, "Antisense oligonucleotides: The next Frontier for treatment of neurological disorders," Nat. Rev. Neurol. **14**, 9–21 (2017).
[13]Y. Krishnan and N. C. Seeman, "Introduction: Nucleic acid nanotechnology," Chem. Rev. **119**, 6271–6272 (2019).
[14]P. W. K. Rothemund, "Folding DNA to create nanoscale shapes and patterns," Nature **440**, 297–302 (2006).
[15]S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, and W. M. Shih, "Self-assembly of DNA into nanoscale three-dimensional shapes," Nature **459**, 414–418 (2009).
[16]B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel, and J. L. Neumann, "A DNA-fuelled molecular machine made of DNA," Nature **406**, 605–608 (2000).
[17]A. J. Turberfield, J. C. Mitchell, B. Yurke, A. P. Mills, M. I. Blakey, and F. C. Simmel, "DNA fuel for free-running nanomachines," Phys. Rev. Lett. **90**, 118102 (2003).
[18]E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf, and J. Kjems, "Self-assembly of a nanoscale DNA box with a controllable lid," Nature **459**, 73–76 (2009).
[19]A.-K. Pumm, W. Engelen, E. Kopperger, J. Isensee, M. Vogt, V. Kozina, M. Kube, M. N. Honemann, E. Bertosin, M. Langecker, R. Golestanian, F. C. Simmel, and H. Dietz, "A DNA origami rotary ratchet motor," Nature **607**, 492–498 (2022).
[20]S. Li, Q. Jiang, S. Liu, Y. Zhang, Y. Tian, C. Song, J. Wang, Y. Zou, G. J. Anderson, J.-Y. Han, Y. Chang, Y. Liu, C. Zhang, L. Chen, G. Zhou, G. Nie, H. Yan, B. Ding, and Y. Zhao, "A DNA nanorobot functions as a cancer therapeutic in response to a molecular trigger *in vivo*," Nat. Biotechnol. **36**, 258–264 (2018).
[21]C. Sigl, E. M. Willner, W. Engelen, J. A. Kretzmann, K. Sachenbacher, A. Liedl, F. Kolbe, F. Wilsch, S. A. Aghvami, U. Protzer, M. F. Hagan, S. Fraden, and H. Dietz, "Programmable icosahedral shell system for virus trapping," Nat. Mater. **20**, 1281–1289 (2021).
[22]Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, and E. Shapiro, "An autonomous molecular computer for logical control of gene expression," Nature **429**, 423–429 (2004).
[23]S. M. Douglas, I. Bachelet, and G. M. Church, "A logic-gated nanorobot for targeted transport of molecular payloads," Science **335**, 831–834 (2012).
[24]A. Chworos, I. Severcan, A. Y. Koyfman, P. Weinkam, E. Oroudjev, H. G. Hansma, and L. Jaeger, "Building programmable jigsaw puzzles with RNA," Science **306**, 2068–2072 (2004).
[25]C. Geary, P. W. K. Rothemund, and E. S. Andersen, "A single-stranded architecture for cotranscriptional folding of RNA nanostructures," Science **345**, 799–804 (2014).
[26]E. K. S. McRae, H. Ø. Rasmussen, J. Liu, A. Bøggild, M. T. A. Nguyen, N. Sampedro Vallina, T. Boesen, J. S. Pedersen, G. Ren, C. Geary, and E. S. Andersen, "Structure, folding and flexibility of co-transcriptional RNA origami," Nat. Nanotechnol. **18**, 808–817 (2023).
[27]L. Zhou, A. R. Chandrasekaran, M. Yan, V. A. Valsangkar, J. I. Feldblyum, J. Sheng, and K. Halvorsen, "A mini DNA–RNA hybrid origami nanobrick," Nanoscale Adv. **3**, 4048–4051 (2021).
[28]X. Wu, Q. Liu, F. Liu, T. Wu, Y. Shang, J. Liu, and B. Ding, "An RNA/DNA hybrid origami-based nanoplatform for efficient gene therapy," Nanoscale **13**, 12848–12853 (2021).
[29]M. F. Parsons, M. F. Allan, S. Li, T. R. Shepherd, S. Ratanalert, K. Zhang, K. M. Pullen, W. Chiu, S. Rouskin, and M. Bathe, "3D RNA-scaffolded wireframe origami," Nat. Commun. **14**, 382 (2023).
[30]A. Marantan and L. Mahadevan, "Mechanics and statistics of the worm-like chain," Am. J. Phys. **86**, 86–94 (2018).
[31]R. Galindo-Murillo and T. E. Cheatham III, "Lessons learned in atomistic simulation of double-stranded DNA: Solvation and salt concerns [Article v1.0]," Living J. Comput. Mol. Sci. **1**, 9974 (2019).
[32]J. Šponer, P. Banáš, P. Jurečka, M. Zgarbová, P. Kührová, M. Havrila, M. Krepl, P. Stadlbauer, and M. Otyepka, "Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome," J. Phys. Chem. Lett. **5**, 1771–1782 (2014).
[33]J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, N. G. Walter, and M. Otyepka, "RNA structural dynamics as captured by molecular simulations: A comprehensive overview," Chem. Rev. **118**, 4177–4338 (2018).
[34]J. Šponer, J. E. Šponer, A. Mládek, P. Banáš, P. Jurečka, and M. Otyepka, "How to understand quantum chemical computations on DNA and RNA systems? A practical guide for non-specialists," Methods **64**, 3–11 (2013).
[35]A. Mládek, M. Krepl, D. Svozil, P. Čech, M. Otyepka, P. Banáš, M. Zgarbová, P. Jurečka, and J. Šponer, "Benchmark quantum-chemical calculations on a complete set of rotameric families of the DNA sugar–phosphate backbone and their comparison with modern density functional theory," Phys. Chem. Chem. Phys. **15**, 7295 (2013).
[36]A. E. Hafner, J. Krausser, and A. Šarić, "Minimal coarse-grained models for molecular self-organisation in biology," Curr. Opin. Struct. Biol. **58**, 43–52 (2019).
[37]S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, "Coarse-grained protein models and their applications," Chem. Rev. **116**, 7898–7936 (2016).
[38]M. DeLuca, S. Sensale, P.-A. Lin, and G. Arya, "Prediction and control in DNA nanotechnology," ACS Appl. Bio Mater. (published online 2023).
[39]T. Sun, V. Minhas, N. Korolev, A. Mirzoev, A. P. Lyubartsev, and L. Nordenskiöld, "Bottom-up coarse-grained modeling of DNA," Front. Mol. Biosci. **8**, 645527 (2021).
[40]N. A. Denesyuk and D. Thirumalai, "Coarse-grained model for predicting RNA folding thermodynamics," J. Phys. Chem. B **117**, 4901–4911 (2013).
[41]R. V. Reshetnikov, A. V. Stolyarova, A. O. Zalevsky, D. Y. Panteleev, G. V. Pavlova, D. V. Klinov, A. V. Golovin, and A. D. Protopopova, "A coarse-grained model for DNA origami," Nucleic Acids Res. **46**, 1102–1112 (2017).
[42]J. Li and S.-J. Chen, "RNA 3D structure prediction using coarse-grained models," Front. Mol. Biosci. **8**, 720937 (2021).
[43]W. K. Dawson, M. Maciejczyk, E. J. Jankowska, and J. M. Bujnicki, "Coarse-grained modeling of RNA 3D structure," Methods **103**, 138–156 (2016).
[44]C. Maffeo, T. T. M. Ngo, T. Ha, and A. Aksimentiev, "A coarse-grained model of unstructured single-stranded DNA derived from atomistic simulation and single-molecule experiment," J. Chem. Theory Comput. **10**, 2891–2896 (2014).
[45]T. E. Cheatham and P. A. Kollman, "Molecular dynamics simulations highlight the structural differences among DNA:DNA, RNA:RNA, and DNA:RNA hybrid duplexes," J. Am. Chem. Soc. **119**, 4805–4825 (1997).
[46]A. Noy, A. Pérez, M. Márquez, F. J. Luque, and M. Orozco, "Structure, recognition properties, and flexibility of the DNA·RNA hybrid," J. Am. Chem. Soc. **127**, 4910–4920 (2005).
[47]J.-H. Liu, K. Xi, X. Zhang, L. Bao, X. Zhang, and Z.-J. Tan, "Structural flexibility of DNA-RNA hybrid duplex: Stretching and twist-stretch coupling," Biophys. J. **117**, 74–86 (2019).

[48]E. de Oliveira Martins, V. Basílio Barbosa, and G. Weber, "DNA/RNA hybrid mesoscopic model shows strong stability dependence with deoxypyrimidine content and stacking interactions similar to RNA/RNA," Chem. Phys. Lett. **715**, 14–19 (2019).
[49]N. Jonoska, N. Obatake, S. Poznanović, C. Price, M. Riehl, and M. Vazquez, "Modeling RNA:DNA hybrids with formal grammars," in *Using Mathematics to Understand Biological Complexity: From Cells to Populations*, edited by R. Segal, B. Shtylla, and S. Sindi (Springer International Publishing, Cham, 2021), pp. 35–54.
[50]E. Poppleton, M. Matthies, D. Mandal, F. Romano, P. Šulc, and L. Rovigatti, "oxDNA: Coarse-grained simulations of nucleic acids made simple," J. Open Source Softw. **8**, 4693 (2023).
[51]T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, "Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model," J. Chem. Phys. **134**, 085101 (2011).
[52]P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. K. Doye, and A. A. Louis, "Sequence-dependent thermodynamics of a coarse-grained DNA model," J. Chem. Phys. **137**, 135101 (2012).
[53]B. E. K. Snodin, F. Randisi, M. Mosayebi, P. Šulc, J. S. Schreck, F. Romano, T. E. Ouldridge, R. Tsukanov, E. Nir, A. A. Louis, and J. P. K. Doye, "Introducing improved structural properties and salt dependence into a coarse-grained model of DNA," J. Chem. Phys. **142**, 234901 (2015).
[54]P. Šulc, F. Romano, T. E. Ouldridge, J. P. K. Doye, and A. A. Louis, "A nucleotide-level coarse-grained model of RNA," J. Chem. Phys. **140**, 235102 (2014).
[55]B. E. K. Snodin, F. Romano, L. Rovigatti, T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, "Direct simulation of the self-assembly of a small DNA origami," ACS Nano **10**, 1724–1737 (2016).
[56]C.-M. Huang, A. Kucinic, J. V. Le, C. E. Castro, and H.-J. Su, "Uncertainty quantification of a DNA origami mechanism using a coarse-grained model and kinematic variance analysis," Nanoscale **11**, 1647–1660 (2019).
[57]E. Benson, A. Mohammed, D. Rayneau-Kirkhope, A. Gådin, P. Orponen, and B. Högberg, "Effects of design choices on the stiffness of wireframe DNA origami structures," ACS Nano **12**, 9291–9299 (2018).
[58]M. C. Engel, D. M. Smith, M. A. Jobst, M. Sajfutdinow, T. Liedl, F. Romano, L. Rovigatti, A. A. Louis, and J. P. K. Doye, "Force-induced unravelling of DNA origami," ACS Nano **12**, 6734–6747 (2018).
[59]E. Torelli, J. W. Kozyra, J.-Y. Gu, U. Stimming, L. Piantanida, K. Voïtchovsky, and N. Krasnogor, "Isothermal folding of a light-up bio-orthogonal RNA origami nanoribbon," Sci. Rep. **8**, 6989 (2018).
[60]B. E. Snodin, J. S. Schreck, F. Romano, A. A. Louis, and J. P. Doye, "Coarse-grained modelling of the structural properties of DNA origami," Nucleic Acids Res. **47**, 1585–1597 (2019).
[61]E. Torelli, J. Kozyra, B. Shirt-Ediss, L. Piantanida, K. Voïtchovsky, and N. Krasnogor, "Cotranscriptional folding of a bio-orthogonal fluorescent scaffolded RNA origami," ACS Synth. Biol. **9**, 1682–1692 (2020).
[62]P. Šulc, T. E. Ouldridge, F. Romano, J. P. Doye, and A. A. Louis, "Modelling toehold-mediated RNA strand displacement," Biophys. J. **108**, 1238–1247 (2015).
[63]N. Srinivas, T. E. Ouldridge, P. Šulc, J. M. Schaeffer, B. Yurke, A. A. Louis, J. P. K. Doye, and E. Winfree, "On the biophysics and kinetics of toehold-mediated DNA strand displacement," Nucleic Acids Res. **41**, 10641–10658 (2013).
[64]F. Romano, D. Chakraborty, J. P. Doye, T. E. Ouldridge, and A. A. Louis, "Coarse-grained simulations of DNA overstretching," J. Chem. Phys. **138**, 085101 (2013).
[65]T. E. Ouldridge, P. Šulc, F. Romano, J. P. Doye, and A. A. Louis, "DNA hybridization kinetics: Zippering, internal displacement and sequence dependence," Nucleic Acids Res. **41**, 8886–8895 (2013).
[66]M. Mosayebi, A. A. Louis, J. P. Doye, and T. E. Ouldridge, "Force-induced rupture of a DNA duplex: From fundamentals to force sensors," ACS Nano **9**, 11993–12003 (2015).
[67]C. Matek, T. E. Ouldridge, J. P. Doye, and A. A. Louis, "Plectoneme tip bubbles: Coupled denaturation and writhing in supercoiled DNA," Sci. Rep. **5**, 7655 (2015).
[68]J. S. Schreck, T. E. Ouldridge, F. Romano, P. Šulc, L. P. Shaw, A. A. Louis, and J. P. Doye, "DNA hairpins destabilize duplexes primarily by promoting melting rather than by inhibiting hybridization," Nucleic Acids Res. **43**, 6181–6190 (2015).
[69]F. Kriegel, C. Matek, T. Geršata, K. Kulenkampff, S. Tschirpke, M. Zacharias, F. Lankaš, and J. Lipfert, "The temperature dependence of the helical twist of DNA," Nucleic Acids Res. **46**, 7998–8009 (2018).
[70]S. K. Nomidis, M. Caraglio, M. Laleman, K. Phillips, E. Skoruppa, and E. Carlon, "Twist-bend coupling, twist waves, and the shape of DNA loops," Phys. Rev. E **100**, 022402 (2019).
[71]A. Suma, V. Carnevale, and C. Micheletti, "Nonequilibrium thermodynamics of DNA nanopore unzipping," Phys. Rev. Lett. **130**, 048101 (2023).
[72]W. Lim, F. Randisi, J. P. K. Doye, and A. A. Louis, "The interplay of supercoiling and thymine dimers in DNA," Nucleic Acids Res. **50**, 2480–2492 (2022).
[73]R. Sharma, J. S. Schreck, F. Romano, A. A. Louis, and J. P. K. Doye, "Characterizing the motion of jointed DNA nanostructures using a coarse-grained model," ACS Nano **11**, 12426–12435 (2017).
[74]Z. Shi, C. E. Castro, and G. Arya, "Conformational dynamics of mechanically compliant DNA nanostructures from coarse-grained molecular dynamics simulations," ACS Nano **11**, 4617–4630 (2017).
[75]W. T. Kaufhold, W. Pfeifer, C. E. Castro, and L. Di Michele, "Probing the mechanical properties of DNA nanostructures with metadynamics," ACS Nano **16**, 8784–8797 (2022).
[76]N. Sugimoto, S. i. Nakano, M. Katoh, A. Matsumura, H. Nakamuta, T. Ohmichi, M. Yoneyama, and M. Sasaki, "Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes," Biochemistry **34**, 11211–11216 (1995).
[77]D. Banerjee, H. Tateishi-Karimata, T. Ohyama, S. Ghosh, T. Endoh, S. Takahashi, and N. Sugimoto, "Improved nearest-neighbor parameters for the stability of RNA/DNA hybrids under a physiological condition," Nucleic Acids Res. **48**, 12042–12054 (2020).
[78]J. SantaLucia and D. Hicks, "The thermodynamics of DNA structural motifs," Annu. Rev. Biophys. Biomol. Struct. **33**, 415–440 (2004).
[79]T. Xia, J. SantaLucia, M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, and D. H. Turner, "Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs," Biochemistry **37**, 14719–14735 (1998).
[80]J. SantaLucia, "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics," Proc. Natl. Acad. Sci. U. S. A. **95**, 1460–1465 (1998).
[81]P. J. A. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, and M. J. L. de Hoon, "Biopython: Freely available Python tools for computational molecular biology and bioinformatics," Bioinformatics **25**, 1422–1423 (2009).
[82]Y. Huang, C. Chen, and I. M. Russu, "Dynamics and stability of individual base pairs in two homologous RNA–DNA hybrids," Biochemistry **48**, 3988–3997 (2009).
[83]S. Wang and E. T. Kool, "Origins of the large differences in stability of DNA and RNA helixes: C-5 methyl and 2′-hydroxyl effects," Biochemistry **34**, 4125–4132 (1995).
[84]S. Whitelam and P. L. Geissler, "Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles," J. Chem. Phys. **127**, 154101 (2007).
[85]G. Torrie and J. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," J. Comput. Phys. **23**, 187–199 (1977).
[86]T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, "Extracting bulk properties of self-assembling systems from small simulations," J. Phys.: Condens. Matter **22**, 104102 (2010).
[87]C. Calladine and H. Drew, "A base-centred explanation of the B-to-A transition in DNA," J. Mol. Biol. **178**, 773–782 (1984).
[88]R. E. Dickerson and H.-L. Ng, "DNA structure from A to B," Proc. Natl. Acad. Sci. U. S. A. **98**, 6986–6988 (2001).
[89]B. Hartmann and R. Lavery, "DNA structural forms," Q. Rev. Biophys. **29**, 309–368 (1996).
[90]N. N. Shaw and D. P. Arya, "Recognition of the unique structure of DNA:RNA hybrids," Biochimie **90**, 1026–1039 (2008).
[91]S. Arnott, R. Chandrasekaran, R. Millane, and H.-S. Park, "DNA-RNA hybrid secondary structures," J. Mol. Biol. **188**, 631–640 (1986).

[92]R. T. Wheelhouse and J. B. Chaires, "Drug binding to DNA·RNA hybrid structures," in *Drug-DNA Interaction Protocols*, edited by K. R. Fox (Humana Press, Totowa, NJ, 2010), pp. 55–70.
[93]J. I. Gyi, A. N. Lane, G. L. Conn, and T. Brown, "Solution structures of DNA·RNA hybrids with purine-rich and pyrimidine-rich strands: Comparison with the homologous DNA and RNA duplexes," Biochemistry **37**, 73–80 (1998).
[94]G. L. Conn, T. Brown, and G. A. Leonard, "The crystal structure of the RNA/DNA hybrid r(GAAGAGAAGC)·d(GCTTCTCTTC) shows significant differences to that found in solution," Nucleic Acids Res. **27**, 555–561 (1999).
[95]J. C. Cofsky, G. J. Knott, C. L. Gee, and J. A. Doudna, "Crystal structure of an RNA/DNA strand exchange junction," PLoS One **17**, e0263547 (2022).
[96]Y. Xiong, "Crystal structure of a DNA·RNA hybrid duplex with a polypurine RNA r(gaagaagag) and a complementary polypyrimidine DNA d(CTCTTCTTC)," Nucleic Acids Res. **28**, 2171–2176 (2000).
[97]A. Marin-Gonzalez, J. G. Vilhena, R. Perez, and F. Moreno-Herrero, "Understanding the mechanical response of double-stranded DNA and RNA under constant stretching forces using all-atom molecular dynamics," Proc. Natl. Acad. Sci. U. S. A. **114**, 7049–7054 (2017).
[98]J. Ji, D. Karna, and H. Mao, "DNA origami nano-mechanics," Chem. Soc. Rev. **50**, 11966–11978 (2021).
[99]M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*, *International Series of Monographs on Physics* (Clarendon Press, Oxford, England, 1988).
[100]T. Odijk, "Stiff chains and filaments under tension," Macromolecules **28**, 7016–7018 (1995).
[101]C. Zhang, H. Fu, Y. Yang, E. Zhou, Z. Tan, H. You, and X. Zhang, "The mechanical properties of RNA-DNA hybrid duplex stretched by magnetic tweezers," Biophys. J. **116**, 196–204 (2019).
[102]D. Y. Zhang and E. Winfree, "Control of DNA strand displacement kinetics using toehold exchange," J. Am. Chem. Soc. **131**, 17303–17314 (2009).
[103]L. Qian and E. Winfree, "Scaling up digital circuit computation with DNA strand displacement cascades," Science **332**, 1196–1201 (2011).
[104]H. Liu, F. Hong, F. Smith, J. Goertz, T. Ouldridge, M. M. Stevens, H. Yan, and P. Šulc, "Kinetics of RNA and RNA:DNA hybrid strand displacement," ACS Synth. Biol. **10**, 3066–3073 (2021).
[105]F. Hong and P. Šulc, "An emergent understanding of strand displacement in RNA biology," J. Struct. Biol. **207**, 241–249 (2019).
[106]R. R. F. Machinek, T. E. Ouldridge, N. E. C. Haley, J. Bath, and A. J. Turberfield, "Programmable energy landscapes for kinetic control of DNA strand displacement," Nat. Commun. **5**, 5324 (2014).
[107]N. E. C. Haley, T. E. Ouldridge, I. Mullor Ruiz, A. Geraldini, A. A. Louis, J. Bath, and A. J. Turberfield, "Design of hidden thermodynamic driving for non-equilibrium systems via mismatch elimination during DNA strand displacement," Nat. Commun. **11**, 2562 (2020).
[108]R. Landgraf, C. B. Chen, and D. S. Sigman, "R-loop stability as a function of RNA structure and size," Nucleic Acids Res. **23**, 3516–3523 (1995).
[109]C. L. Lawson, A. Patwardhan, M. L. Baker, C. Hryc, E. S. Garcia, B. P. Hudson, I. Lagerstedt, S. J. Ludtke, G. Pintilie, R. Sala, J. D. Westbrook, H. M. Berman, G. J. Kleywegt, and W. Chiu, "EMDataBank unified data resource for 3DEM," Nucleic Acids Res. **44**, D396–D403 (2015).
[110]E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, and T. E. Ferrin, "UCSF ChimeraX: Structure visualization for researchers, educators, and developers," Protein Sci. **30**, 70–82 (2020).
[111]J. Bohlin, M. Matthies, E. Poppleton, J. Procyk, A. Mallya, H. Yan, and P. Šulc, "Design and simulation of DNA, RNA and hybrid protein–nucleic acid nanostructures with oxView," Nat. Protoc. **17**, 1762–1788 (2022).