ANNaMo: Coarse-grained modelling for folding and assembly of RNA and DNA systems

F. Tosti Guerra,[1,a)] E. Poppleton,[2,3] P. Šulc,[2,4] and L. Rovigatti[1,b)]
[1)]*Department of Physics, Sapienza University of Rome*
[2)]*School of Molecular Sciences and Center for Molecular Design and Biomimetics, The Biodesign Institute, Arizona State University*
[3)]*Biophysical Engineering Group, Max Planck Institute for Medical Research*
[4)]*Department of Bioscience, School of Natural Sciences, Technical University Munich*

(Dated: May 2, 2024)

The folding of RNA and DNA strands plays crucial roles in biological systems and bionanotechnology. However, studying these processes with high-resolution numerical models is beyond current computational capabilities due to the timescales and system sizes involved. In this article, we present a new coarse-grained model for investigating the folding dynamics of nucleic acids. Our model represents 3 nucleotides with a patchy particle and is parametrized using well-established nearest-neighbor models. Thanks to the reduction of degrees of freedom and to a bond-swapping mechanism, our model allows for simulations at timescales and length scales that are currently inaccessible to more detailed models. To validate the performance of our model, we conducted extensive simulations of various systems: We examined the thermodynamics of DNA hairpins, capturing their stability and structural transitions, the folding of an MMTV pseudoknot, a complex RNA structure involved in viral replication, and also explored the folding of an RNA tile containing a k-type pseudoknot. Finally, we evaluated the performance of the new model in reproducing the melting temperatures of oligomers and the dependence on the toehold length of the displacement rate in toehold-mediated displacement processes, a key reaction used in molecular computing. All in all, the successful reproduction of experimental data and favorable comparisons with existing coarse-grained models validate the effectiveness of the new model.

I. INTRODUCTION

RNA and DNA molecules play critical roles in biological systems and have become increasingly important in constructing nanoscale architectures[1,2]. Successful bionanotechnology designs include DNA origami[3], a nanostructure consisting of a long (about 7000 bases) scaffold single-stranded DNA strand, which is compacted into the target shape by designed shorter staple strands that connect different regions on the scaffold. More recently, DNA bricks[4] consisting just of short strands have been shown to assemble both into 2D and 3D target shapes. Finally, single-stranded DNA and RNA origami and tile structures[5–7] have been designed and shown to be able to also fold into 2D or 3D shapes. These designs consist of a single strand of length ranging from hundreds to thousands of bases, where different regions are complementary to each other, designed to fold into a compact structure consisting of duplex regions and crossovers. DNA and RNA nanostructures have found a range of applications, ranging from material science to biomedicine[2].

The nucleic acid nanostructure designs are, for the most part, based on DNA and RNA thermodynamics, where the target design maximizes the number of Watson-Crick (and wobble) base pairs present in the system. However, understanding the kinetics of the folding processes of these designs is essential for unraveling their assembly mechanism and optimizing yields, since kinetic traps can prevent access to the conformations corresponding to global free-energy minima. Access to effective simulation of the folding pathway will facilitate the design of new and more complex nanoscale structures, with applications ranging from nanomanufacturing of plasmonic and photonic devices, molecular robotics, and computing, to the design of more sophisticated tools for biomedical diagnostics and therapeutics.

An additional area of interest is the folding of nucleic acids in biological systems, such as single-stranded genomes of viruses, contranscriptional folding of nascent RNA[8], as well as folding of functional RNA molecules, such as designed mRNA for optimized vaccine applications[9]. Understanding folding pathways enables the rational design of functional RNA structures with desired properties, such as enhanced catalytic activity or improved binding affinity to specific targets[10,11].

Computational models have been developed to explore the processes involved in RNA and DNA folding. These models range from atomistic to coarse-grained resolutions, each offering unique advantages and insights into the dynamics of folding. The all-atom model is a commonly employed approach that explicitly represents individual atoms and their interactions. Atomistic force fields such as AMBER[12] and CHARMM[13] have been extensively used to investigate the folding of nucleic acids[14,15]. While these models accurately capture atomistic details and provide valuable insights into the structural and energetic aspects of RNA and DNA, their simulations are computationally demanding and offer limited access to the timescales of interest.

To overcome the limitations of all-atom models, coarse-

[a)]francesco.tostiguerra@uniroma1.it
[b)]lorenzo.rovigatti@uniroma1.it

grained models, which simplify the representation of nucleic acids by grouping multiple atoms into a single particle or bead, have emerged as powerful tools for studying longer lengths and timescales. Several coarse-grained models have been specifically developed for studying DNA and RNA[16–37].

Coarse-grained models typically treat the solvent and solvated ions implicitly and represent groups of atoms in the DNA/RNA structure with effective interactions. This simplification enables the study of larger and more extended molecular systems for longer times. Coarse-graining techniques can be broadly categorized as either bottom-up or top-down, each aiming to capture specific aspects of the system. The bottom-up approach formally maps the statistical behavior of a more detailed model into a coarse-grained description, while the top-down approach aims to reproduce as many experimentally relevant properties as possible[38].

Many of the models mentioned above are parametrized using nearest-neighbor (NN) models. Initially introduced by Poland and Scheraga to investigate duplex denaturation phase transitions[39], this approach has been meticulously developed in subsequent years[40–49] to describe binding equilibria for oligonucleotides. Nearest-neighbor models calculate the free energy change (ΔG) associated with forming a duplex by summing the contributions of individual base pair interactions. These interactions are characterized by experimentally determined enthalpy (ΔH_{ij}) and entropy (ΔS_{ij}) values for each possible combination of adjacent base pairs (i, j). Different base pair combinations have different thermodynamic parameters, reflecting variations in hydrogen bonding, stacking interactions, and other factors.

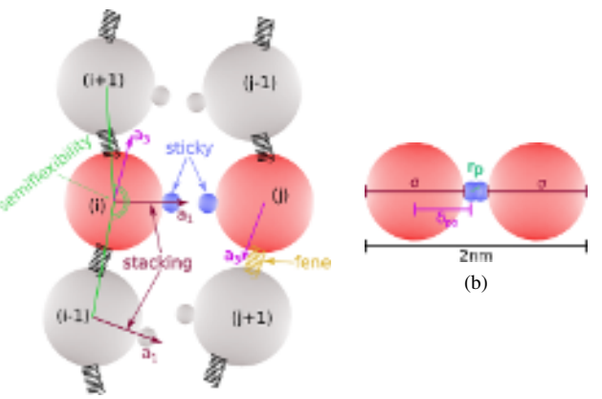
In a NN framework, the binding equilibrium between two isolated strands A and B and their associated fully-bound duplex product AB is fully characterised by the equilibrium constant (K), which is defined as the ratio of the concentration of the product $[AB]$ to the concentrations of the individual strands $[A]$ and $[B]$:

$$K = \frac{[AB]}{[A][B]} \propto \exp\left(-\beta(\Delta H - T\Delta S)\right), \tag{1}$$

where ΔH and ΔS are the total enthalpy and entropy change upon binding, and $\beta = 1/k_B T$. By considering nearest-neighbor interactions along the entire sequence length, the model predicts the stability of duplexes and provides insights into their melting temperatures, binding affinities, and overall thermodynamic properties.

In our study, we introduce a new coarse-grained description of nucleic acids where a single patchy particle represents, in principle, n nucleotides, enabling efficient exploration of RNA and DNA folding processes. We parametrize the model using well-established nearest-neighbor models[40–45,49] for DNA and RNA thermodynamics, the oxDNA coarse-grained model[50], as well as estimates of DNA persistence lengths[51,52], to capture essential interactions involved in folding dynamics while maintaining computational efficiency. The new model, named ANNaMo (Another Nucleic-acid Nanotechnology Model), combines the strengths of existing coarse-grained and patchy particle models, providing new possibilities for investigating folding phenomena at previously inaccessible time- and length-scales. This first iteration of the model has been parametrized by fixing $n = 3$, and its thermodynamic performance has been compared with available experimental and numerical data.

II. MODEL DESCRIPTION



(a) (b)

Figure 1. (a) schematic representation of the interaction potentials acting between beads in ANNaMo (b) the diameter of the DNA/RNA helix is approximately 2 nm. Here the bead diameter is σ, and since we set the distance between the patch and the bead surface to $\delta_{pb} = 0.65\sigma$ and the patch diameter to $r_p = 0.219\sigma$, we can determine that $2\sigma + 2\sigma(\delta_{pb} - 0.5) + r_p\sigma = 2$ nm. Therefore, we can deduce that $\sigma = 2/(1 + 2\delta_{pb} + r_p)$, or 0.79 nm.

In ANNaMo, each DNA/RNA strand is modelled as a polymer composed of N beads with diameter σ. Each bead is decorated with an attractive spherical patch, having a radius of 0.219σ and positioned $\delta_{pb} = 0.65\sigma$ away from the center of the bead, and represents n consecutive nitrogenous bases. We incorporate several potentials to reproduce the thermodynamic and mechanical characteristics of these biopolymers.

Figure 1 provides a visual summary of the notation and different interaction terms used in our model, while the detailed functional forms are reported in Appendix A. The potentials between topologically bonded beads linked through the backbone include the Kremer-Grest force field[53], combining a Weeks-Chandler-Andersen (WCA) potential to model excluded volume, and a finitely extensible nonlinear elastic (FENE) potential to mimic the covalent bonds along the strand. In addition, we introduce a term to add stiffness to the structure: a three-body potential (V_{semiflex}) that tends to align three consecutive beads, resulting in a different persistence length between single and double strands (see II C). To model the stacking of bases in DNA/RNA, we incorporate a term called V_{stack}, which promotes the alignment of the directions of consecutive beads that determine the patch positions by using a cosine-angle potential.

For the non-bonded interactions, we used the WCA potential to account for excluded-volume interactions. Additionally, we employ a patch-patch interaction potential, based on

the functional form proposed by Stillinger and Weber[54], to model the hybridization of nucleotides. The strength of the attraction between two beads i and j, ε_{ij}, depends on the types of nucleotides described by each interacting bead and is computed with a nearest-neighbor (NN) model as described below. To ensure that each bead can bind to only one other bead, we implement a repulsive three-body interaction (V_{3b}) that penalizes the formation of triplets of bonded beads[55]. This repulsive potential compensates for the gain associated with the formation of a second bond and can be tuned to favor bond swapping. More, specifically, the parameter λ in V_{3b} allows to interpolate between the limits of swapping ($\lambda = 1$) and non-swapping ($\lambda \gg 1$) bonds. Finally, the V_{sticky} potential is modulated by a term V_{direct} that takes into account the $\vec{a}_3$ orientations of the beads to ensure that only antiparallel strands can bind to each other. We note that one of the consequences of the coarse nature of the model is that, differently from real nucleic acids[56], single- and double-stranded chains modelled with ANNaMo have similar contour lengths (as shown by the bead-bead distance, which is the same for single- and double-strands, see Appendix D).

In the following, all ANNaMo molecular dynamics simulations are run at constant internal temperature $T_i = 1$[57] using the open-source oxDNA simulation engine[58]. Note that T_i is different from the actual physical temperature T that we vary in order to evaluate melting temperature and that controls the strength of the bead-bead attraction, as described in the section below. The equations of motion are integrated with a velocity Verlet algorithm with a time step $\Delta t = 0.002$ (in internal units), a value that has been chosen by checking the energy conservation in microcanonical test simulations. A snapshot of the code used and examples linked to the systems investigated in this work are freely available online[59]. Most of the simulations were run for 10^9 steps, which took a few hours on a single CPU core, while the tile was simulated for 5×10^9 steps, which took around 30 hours on a single core. In addition to ANNaMo simulations, we also run some oxDNA2[50] simulations to evaluate the melting curves of DNA hairpins through umbrella sampling simulations exploiting the Virtual Move Monte Carlo algorithm[38]. Each hairpin required 5 single-core simulations at different temperatures that lasted for approximately 2 days each.

A. Bead design

The loss of resolution caused by dividing a strand into beads of n nucleotides leads to a problem where native motifs (*e.g.* hairpins in natural RNAs or crossovers in origami structures) may contain nucleotides from multiple beads or a motif may occur in the middle of a bead. We do not want, however, to make each bead correspond to a unique interaction, thereby losing the ability to simulate unintended folds caused by competing complementary sequences. To balance these demands, our bead design is based on the native contacts observed in folded or designed structures, with the aim of having the majority of the beads representing the same number of nucleotides, n. The optimal number of nucleotides per bead is, in general, determined by the desired level of detail and the specific system at hand. Here we optimise the model parameters for $n = 3$, which in the following will be used as the standard size of the beads. However, we allow for a small fraction of the remaining beads to deviate by one nucleotide, either more or fewer, from the standard length, n. As a consequence, the beads are composed of varying numbers of nucleotides, requiring the establishment of a rule for calculating interactions. Specifically, we use the strongest ΔG found by pairing the shorter bead with all possible subsequences of the same length contained within the longer bead.

In the future we will extend the model to make it possible to simulate DNA and RNA structures at varying levels of detail by providing n-dependent sets of parameters in the future.

B. Tuning of the bead-bead attraction strength

The strength of the sticky interaction between any two beads i, j is controlled by a term $\beta\varepsilon_{ij}$ that depends on the nucleotide sequences present in each bead. This term models the pairing of the nucleotides that compose the two beads by taking into account the standard-state Gibbs free energy, multiplied by β, as provided by the nearest-neighbor model. In the following, we will consider systems simulated at a fixed monovalent salt concentration of 0.5 M, but any other condition can be considered, as long as the NN model's parameters support it.

In order to properly account for temperature variations, we need to separate the contributions from enthalpy and entropy. This can be achieved by rewriting $-\Delta G^\circ/k_BT$ as follows:

$$\begin{aligned} -\frac{\Delta G^\circ}{k_BT} &= -\left(\frac{\Delta H^\circ}{k_BT} - \frac{\Delta S^\circ}{k_B}\right) \\ &= -\left(\frac{\Delta H^\circ}{k_BT_{\text{ref}}}\frac{T_{\text{ref}}}{T} - \frac{\Delta S^\circ}{k_B}\right) \end{aligned} \tag{2}$$

where T_{ref}/T quantifies the temperature difference between the simulation temperature and the temperature T_{ref} used to estimate the values of $\Delta H^\circ, \Delta S^\circ, \Delta G^\circ$ as reported in NN models. Typically, as it is the case here, $T_{\text{ref}} = 37\ ^\circ\text{C}$.

In NN models, the stability of each base pair is influenced by its nearest neighbours (hence the name). Here we incorporate this effect by taking into account the nucleotide base steps at the boundaries, *i.e.* the last nucleotide of the previous bead and the first nucleotide of the following bead, for each pair of beads involved in the interaction calculation. However, in order to obtain the correct free-energy difference for duplex formation, the contributions due to nucleotides outside of the considered beads, which are counted twice, are halved. As an example, consider two sequences: $5' - (..A)(CGG)(C..).....((..G)(CCG)(U..) - 3'$ (see Figure 2). The ΔH° and ΔS° for the interaction between the two mid-

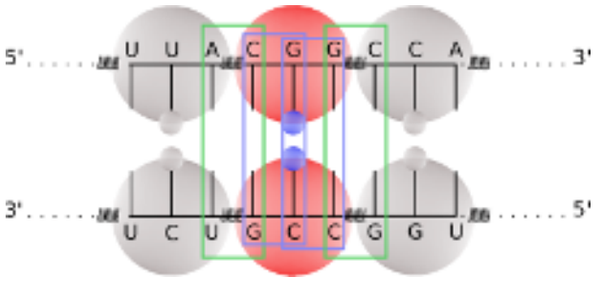


Figure 2. Contributions to the interaction strength between the two colored (red) beads, illustrating the base steps involved. The blue boxes surround base steps made of nucleotides within the beads, while the green boxes surround the two boundary base steps (*i.e.* base steps comprising nucleotides both inside and outside of the considered beads).

dle beads will have the following contributions:

$$\begin{pmatrix} 5'-A)(CGG)(C-3' \\ 3'-U)(GCC)(G-5' \end{pmatrix} = \frac{1}{2}\begin{pmatrix} AC \\ UG \end{pmatrix} + \begin{pmatrix} CG \\ GC \end{pmatrix} + \begin{pmatrix} GG \\ CC \end{pmatrix} + \frac{1}{2}\begin{pmatrix} GC \\ CG \end{pmatrix}. \quad (3)$$

In order to establish a meaningful correspondence between the potential used in our simulations and the NN parameters, the attraction strength between two beads i and j, ε_{ij}, is computed by equaling the free-energy change due to a bond with that predicted by the NN model, $\Delta H - T\Delta S$[60], which for the form and parameters of the bead-bead attraction used here gives

$$\beta\varepsilon_{ij} = -\frac{1}{\alpha}\left[\left(\frac{\Delta H^{\ominus}}{k_B T_{37}}\frac{T_{37}}{T} - \frac{\Delta S^{\ominus}}{k_B}\right) + \ln\left(\rho^{\ominus}\sigma^3\tilde{V}_b\right)\right], \quad (4)$$

where $\rho^{\ominus} = [\ominus]\mathcal{N}_A = 6.022\times10^{26}\,\mathrm{m}^{-3}$ is the standard number density, σ is the diameter of the beads and corresponds to 0.79 nm (see Figure 1b), while $\tilde{V}_b = 0.0019$ and $\alpha = 0.89$ are computed numerically (see Appendix B for additional details).

We note that ε_{ij} is the only parameter affected by the type and nature of the nucleotides: in ANNaMo the geometry and structure of the strands are otherwise independent of the nucleic acid modeled (DNA or RNA) and bead sequence.

We also include the $\Delta H^{\ominus}$ and $\Delta S^{\ominus}$ associated with the initiation and the terminal penalties estimated by NN models[40,49]. The implementation of the terminal correction is straightforward: given two terminal beads, if the first bead initiates a base step with the last two nucleotides of the second bead, forming a base step tabulated in NN models, we include the associated $\Delta H^{\ominus}_{\mathrm{terminal}}$ and $\Delta S^{\ominus}_{\mathrm{terminal}}$. By contrast, the initiation term cannot be assigned to two specific beads for each pair of strands, since it is not known beforehand which pair of beads will bond first. As a result, we distribute the initiation contribution evenly among all potential duplex beads: when computing the strength of the interaction between two beads belonging to different strands, we add to it $\Delta H^{\ominus}_{\mathrm{init}}$ and $\Delta S^{\ominus}_{\mathrm{init}}$, as given by NN models, divided by the length (expressed in number of beads) of the shorter strand. In this way, the total free-energy difference between each pair of strands is always equal to that estimated by the NN model, without having to resort to ad-hoc simulation procedures (where *e.g.* the attraction strength between two beads depends on the simulation state and it is therefore time-dependent).

C. Tuning of the parameters

In addition to the bead-bead attraction strengths, ANNaMo features three more free parameters: the strength, denoted as k_s, and width parameter, ξ, of the V_{semiflex} potential, as well as the stacking strength, η, of the V_{stacking} potential. We fit the values of these quantities by optimising the melting temperatures of DNA hairpins with respect to the predictions of the empirically-parameterized oxDNA model, and to the persistence lengths of single- and double-stranded DNA, making ANNaMo a top-down model. For the thermodynamic data, we focus on two sets of hairpins: twelve hairpins with a stem length of six nt and a loop length of six nt, and twelve hairpins with a stem length of six nt and a loop length of nine nt. We keep the loop sequence constant, while each hairpin's stem is randomly generated. The melting temperature is defined as the temperature at which the yield of the closed hairpin is 0.5, where a hairpin is considered closed in oxDNA simulations when at least 3 nucleotides are bonded, and for ANNaMo when at least one bead-bead bond is present. The estimation of the persistence lengths is carried out by simulating 200-bead hairpins with stems of 98 beads, and 100-bead single strands. In both cases, we define the unit vector $\hat{v}_i$ connecting each pair i and $i+1$, and define $\cos(\theta_{ij}) = \hat{v}_i \cdot \hat{v}_j$, where i and j are separated by m beads. Averaging over all pairs (excluding the terminal 10 beads), we obtain an angular correlation, $\langle\cos(\theta_{ij})\rangle$, which is a function of m only. For worm-like chains, such a correlation decays as

$$\langle\cos(\theta_{ij})\rangle = e^{-m/l_p}. \quad (5)$$

We apply Eq. (5) to extract the persistence length from our simulations. While the potential we use yields persistence lengths that are sensibly different between single- and double-stranded molecules, it also overestimates the presence of kinks in double strands, which decreases their persistence length. Therefore, we perform the analysis described above with and without kinks, which are defined as i,j neighboring pairs for which $V_{\mathrm{semiflex}}(\theta_{ij}) - k_s > -k_B T_i$ (see also Appendix E). While the proper persistence length requires that all angles are taken into account, excluding kinks provides an estimate of the rigidity of double-stranded parts that are mechanically stabilised by neighbouring motifs of the structure they are part of, which are common in complex nanostructures such as origami. Therefore, we argue that the two values of the persistence length (with and without kinks) provide an estimate of the range of the rigidity of double strands modelled with ANNaMo.

Through extensive analysis, we found that the closest results to the oxDNA predictions and the estimated persistence

length for single- and double-stranded DNA are obtained when using $\beta k_s = 4$, $\xi = 0.07$ and $\beta\eta = 6$. The resulting melting temperatures are shown in Table I, comparing very well to those extracted from oxDNA simulations. Indeed, the two data sets differ by an average of just about two kelvins.

Figure 3 illustrates the persistence lengths of double and single strands expressed in number of beads. Since our model does not distinguish between ss- and ds-chain we used the same conversion rule to express the persistence length in nm (refer to Appendix D). The persistence length of the double strand is approximately 27.7 beads with kinks and 65.9 without kinks, equivalent to around 21.2 and 50.5 nm. In contrast, the derived persistence length of single strands is approximately 3 beads, corresponding to 2.3 nm. These values align well with estimates for real DNA, where the persistence length of double-stranded DNA is approximately 50 nm[51], and that of single-stranded DNA is approximately (1.98 ± 0.72) nm[52].

6nt stem 6nt loop	oxDNA	ANNaMo	Δ
h1: GCGTTGCTTCTCCAACGC	66	67	1
h2: TTGGCGCTTCTCCGCCAA	72	72	0
h3: AGGCTCCTTCTCGAGCCT	64	66	2
h4: CTCAGGCTTCTCCCTGAG	61	59	-2
h5: GGACGTCTTCTCACGTCC	64	67	3
h6: CGTGGACTTCTCTCCACG	64	62	-2
h7: CGCAACCTCTTCGTTGCG	68	64	-4
h8: AACCGCCTCTTCGCGGTT	71	71	0
h9: TCCGAGCTCTTCCTCGGA	63	66	3
h10: GAGTCCCTCTTCGGACTC	60	59	-1
h11: CCTGCACTCTTCTGCAGG	65	64	-1
h12: GCACCTCTCTTCAGGTGC	65	68	3
			$\langle\lvert\Delta\rvert\rangle$
			1.8

6nt stem 9nt loop	oxDNA	ANNaMo	Δ
h1: GCGTTGCTATGCTTCCAACGC	62	65	3
h2: TTGGCGCTATGCTTCCGCCAA	66	68	2
h3: AGGCTCCTATGCTTCGAGCCT	60	62	2
h4: CTCAGGCTATGCTTCCCTGAG	57	56	-1
h5: GGACGTCTATGCTTCACGTCC	60	64	4
h6: CGTGGACTATGCTTCTCCACG	59	60	1
h7: CGCAACCTACGTTTCGTTGCG	61	62	1
h8: AACCGCCTACGTTTCGCGGTT	65	67	2
h9: TCCGAGCTACGTTTCCTCGGA	59	63	4
h10: GAGTCCCTACGTTTCGGACTC	56	56	0
h11: CCTGCACTACGTTTCTGCAGG	59	62	3
h12: GCACCTCTACGTTTCAGGTGC	60	65	5
			$\langle\lvert\Delta\rvert\rangle$
			2.3

Table I. Melting temperatures of selected hairpins (in °C), as obtained with oxDNA and ANNaMo. $|\Delta|$ is the absolute value of the difference between the oxDNA and ANNaMo predictions, and $\langle|\Delta|\rangle$ is its average.

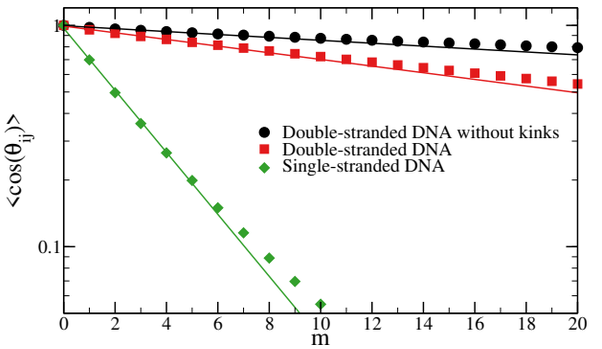


Figure 3. Angular correlation as a function of the chemical distance m for double-stranded and single-stranded DNA molecules. Points are simulation data, lines are exponential fits performed in the $m \in [0,5]$ interval. Note that, as for real DNA and RNA, single strands do not behave as worm-like chains and therefore the angular correlation decays only approximately as an exponential.

III. RESULTS

A. DNA hairpins

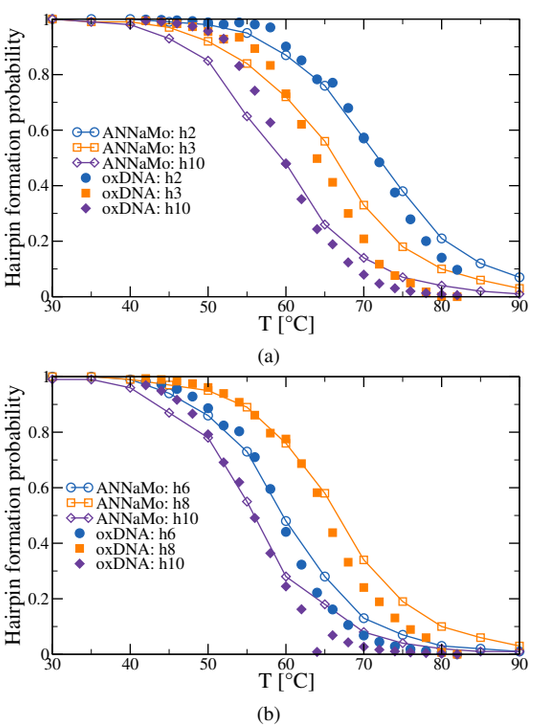


Figure 4. Comparison of the melting curves predicted by oxDNA (points) and ANNaMo (lines) for selected DNA hairpins with a stem length of six nt and a loop length of six nt (a), and selected DNA hairpins with a stem length of six nt and a loop length of nine nt (b). The resulting melting temperatures are listed in Table I.

We start by comparing the melting curves of DNA hairpins simulated with oxDNA and ANNaMo. We use the same def-

inition of a closed hairpin as in Section II C. A representative set of these results, taken from the dataset used to tune the model parameters (see Table I), is shown in Figure 4. For all the hairpins considered, the agreement is almost quantitative for a large temperature range, with oxDNA hairpins exhibiting a somewhat narrower transition, *i.e.* slightly steeper melting curves. At high temperatures the ANNaMo data seems to fall off more slowly compared to oxDNA, demonstrating a slight overestimation of the stability of the double-stranded parts. We ascribe this behavior to the fact that the bonding volume of bead-bead interactions is rather large: even at very high temperature, where the attraction is negligible, we find a sizeable number of configurations in which two compatible beads are mutually positioned and oriented such that they result bonded, especially for small chains like the ones investigated here.

We use hairpin systems to also estimate the performance difference between oxDNA and ANNaMo. Specifically, we run single-core molecular dynamics simulations of hairpins at the melting temperature and count the number of transitions between the two states (open and closed). Dividing the resulting number by the simulation wall time yields a factor of ≈ 100, which we use to estimate the sampling speed-up of the new model compared to oxDNA. While the precise figure should (and will) depend on the system at hand, it is reasonable to expect the speed-up to be of the same order of magnitude, independently on the specific conditions.

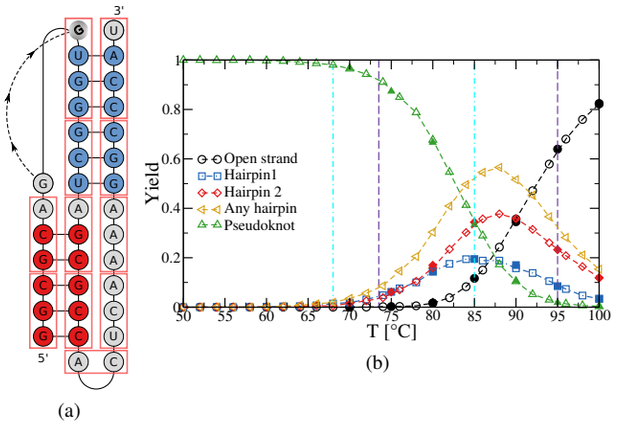


Figure 5. (a) A schematic representation of the secondary structure of the MMTV pseudoknot, showing the bead division used to simulate the strand with ANNaMo. Nucleotides inside the same box belong to the same bead, while nucleotides outside the box belong to the bead that follows. (b) ANNaMo equilibrium yields for the MMTV pseudoknot. Filled symbols represent the results of simulations without swapping, while empty symbols represent those with swapping. The transition temperatures determined experimentally[65] and numerically, with oxRNA[26], are indicated by violet dashed lines and cyan dash-dotted lines, respectively.

B. RNA pseudoknot

A pseudoknot is a folding pattern that involves the formation of additional base pairs between distant regions of the RNA sequence, resulting in a knot-like structure. Pseudoknots are known to have crucial functions in various biological processes such as regulation of gene expression and ribosome function[61], and viral replication[62]. They have also been proposed to have a notable influence on RNA folding pathways[63]. The thermodynamics of pseudoknot base-pairing remains poorly characterized due to their enormous diversity and complexity. Furthermore, the fact that they are composed of non-nested base pairs means that they cannot be computed using the efficient dynamic-programming approaches which dominate in the RNA secondary structure prediction field[64]. Therefore, exact and heuristic coarse-graining methods have mainly focused on pseudoknot-free secondary structures.

We test our model by exploring the melting curve of the well-known MMTV pseudoknot[65]. We divide the sequence in beads as depicted in Figure 5a and we compare our predictions with experimental calorimetry measurements[65] and oxRNA simulations[26], in Figure 5b. The yields of the two pseudoknot-precursor motifs (hairpin 1 and 2, highlighted in blue and red in Figure 5a) display the same qualitative trends observed in oxRNA (reported in Ref.[26]), although the peaks for hairpin 1 and hairpin 2 are higher and lower in ANNaMo compared to oxRNA, respectively. Summing up the yields of the two intermediate hairpins we obtain a curve (gold points in the figure) whose intersection with the open strand and pseudoknot data identifies the two transition temperatures, 83 ° C and 92 °C. Compared to the experimental temperatures (74 °C and 95 °C), these values are at least as good as the ones obtained with oxRNA simulations (68 °C and 85 °C) which, given the lower resolution and much higher computational efficiency of the ANNaMo compare to the latter, is an excellent result.

We have also run simulations without the swap mechanism, which when enabled serves the purpose of speeding up equilibration and sampling, by setting $\lambda = 10$. As shown in Figure 5b, the results obtained with and without the swap (empty and full symbols, respectively) overlap perfectly, demonstrating that the thermodynamics is not affected by the value of λ.

C. RNA tile

Drawing inspiration from the idea of molecular tiles, which are individual units that self-assemble to create intricate patterns or structures[66], we simulated the assembly of a single-stranded RNA tile containing a k-type pseudoknot[67].

Unfortunately, no comparison with oxRNA is possible, as observing the folding of such a structure, composed of 132 nucleotides, is currently out of reach from the computational point of view. As before, we split the sequence into beads of average size 3, but optimize the division by hand so that nucleotides that are supposed to be paired in the native structure belong to the same beads.

We simulate the tile for different temperatures, evaluating the fraction of bonds between any two beads, and between beads that should be bonded in the native structure. The results are shown in Figure 6. The ANNaMo model predicts a

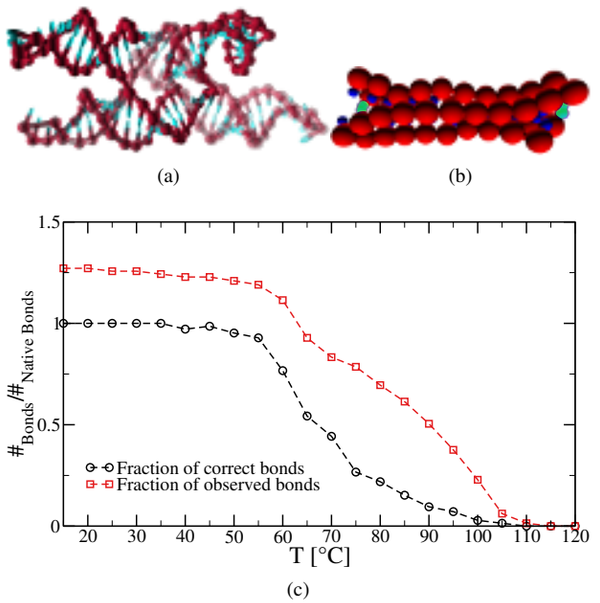


Figure 6. (a) The RNA tile composed of 132 nucleotides[67] we simulate, as represented by oxRNA. (b) A ANNaMo configuration simulated at 28°C. The highlighted (green) patches form kissing loops.(c) Melting curve of the RNA tile simulated with the ANNaMo model.

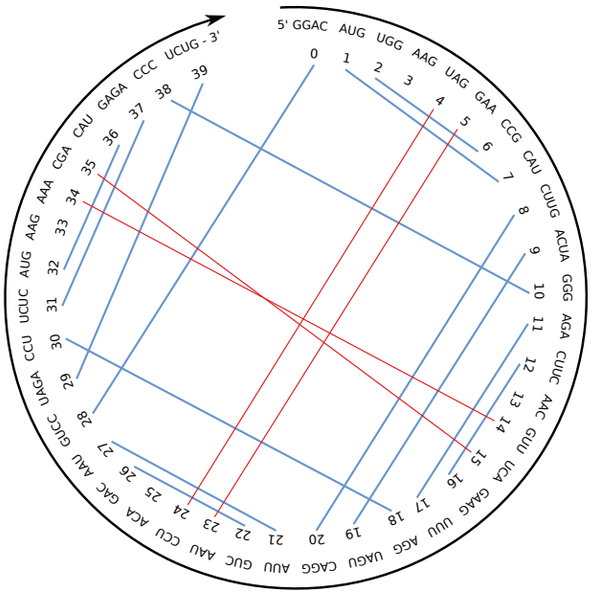


Figure 7. Circular representation of the RNA tile structure and of its sequence, alongside the adopted division into beads (numbers) and the bonds that we observe at low temperature (lines). Blue lines indicate native bonds, while red lines indicate the most probable bonds that connect beads that are not fully complementary but are nevertheless observed in simulations.

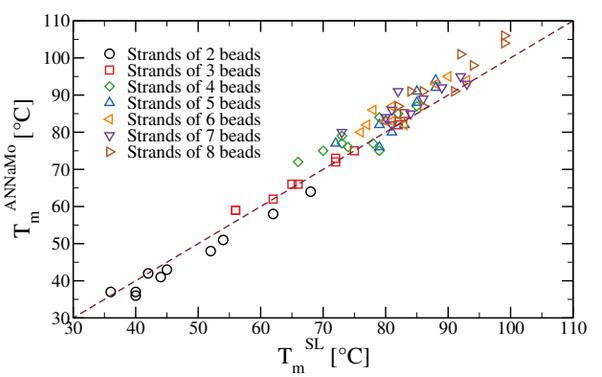


Figure 8. Melting temperatures (T_m) predicted by ANNaMo as a function of the corresponding predictions of the SantaLucia model for 70 duplexes of 6 to 24 bp in length. The average absolute deviation is ≈ 3 °C.

melting temperature (*i.e.* a temperature at which half of the native contacts are formed) around 65°, and that below ≈ 50° nearly all bonds are formed.

At low temperatures, on top of the native contacts, which are always formed, we observe additional bonds that link beads that are not fully compatible with each other. Figure 7 shows the tile sequence, the splitting into beads and the native contacts (blue lines), as well as the most probable misbonds observed (red lines). It is clear that most misbonds happen between beads that are part of hairpin loops, thereby forming so-called kissing loops (highlighted in figure 6b). Kissing loops complexes are known to play an important role in RNA-RNA interactions, both in the biological and nanotechnology contexts[68–70]. Although it is hard to estimate the stability of these motifs within the tile with oxRNA or similar coarse-grained models, the tile arms of the real system should be rigid enough to disfavour these particular kissing loops. Therefore, it is possible that the ANNaMo model overestimates their stability.

D. DNA double strands

Without further parameter optimizations, we compared the melting temperatures of DNA oligomers with that predicted by the SantaLucia's NN model. We simulated duplexes of varying lengths, ranging from 2 to 8 beads made of 3 nucleotides each, equivalent to 6 to 24 base pairs (refer to Table II for detailed DNA sequences and the corresponding simulation box sizes utilized in the study). As we simulate small systems where a single assembly can form, in order to mitigate finite-size effects resulting from the suppression of concentration fluctuations, we employed the approach introduced in Ref.[71].

The results, reported in Figure 8, show that the ANNaMo melting temperatures are always rather close to those predicted by NN models, with an average absolute deviation of ≈ 3 °C, comparable to that observed for the hairpins investigated earlier. The model tends to underestimate melting temperatures of short strands, and overestimate those of longer strands, with oligomers of length 3 and 4 beads (9-12 bp), which is a range of common domain sizes in DNA nanotech-

nology, performing best.

E. Toehold-mediated strand displacement

Toehold-mediated strand displacement (TMSD) processes are a key mechanism used in molecular computing[72]. It consists of an invader strand that displaces an incumbent strand that was previously bound to a substrate strand. The substrate can be longer than the incumbent strand, resulting in a single-stranded region (toehold) to which the invading strand can bind. The length of the toehold can be used to tune the kinetics of DNA- and RNA-nanotechnology systems[73]. Especially in larger reaction cascades consisting of many strand species, using the toehold length to fine-tune the kinetics of the individual strand-displacement reactions can be required to achieve good performance of the molecular circuit[74]. The biophysics of the process has been investigated in detail with both experiments and coarse-grained simulations[75–77]. However, most off-lattice 3D coarse-grained models are too detailed to directly probe TMSD events in an unbiased fashion, and rare-event techniques, such as umbrella sampling or forward-flux sampling (FFS)[78,79], have to be deployed in order to obtain reliable estimates of thermodynamic and kinetic quantities.

Here we use the three-strand system studied in Refs.[75,76]: a 20-nucleotide-long incumbent strand is complementary (and bound) to a substrate that has an additional toehold of variable length (ranging from 0 to 15 nucleotides); the third strand is perfectly complementary to the substrate. If the toehold is present, then the stable thermodynamic state is the one where the invader is bound to the substrate after having displaced the incumbent. If no toehold is present, the incumbent-substrate and invader-substrate states have the same free energy. The main kinetic quantity of interest is the displacement rate, which can be estimated through experiments, as well as through numerical simulations of coarse-grained models[75–77].

Here we use ANNaMo to evaluate the displacement rate as a function of toehold length (up to 5 beads, corresponding to 15 nucleotides) with and without the bond-swapping mechanism. As done in Ref.[76] for oxDNA, we use an interaction matrix where the only non-zero entries are those relative to the native contacts, and we do not take into account the time spent in three-stranded complexes to evaluate rates in order to make it possible to compare results with experiments. For the swapping case and for the non-swapping cases with non-zero toehold length, the dynamics of ANNaMo is fast enough that brute-force calculations are possible. In the other cases we resort to performing FFS calculations.

Figure 9 shows the relative displacement rates as a function of toehold length as obtained with ANNaMo and compared with experiments and oxDNA simulations (Fig. 9a), as well as a comparison between the relative displacement rates with and without bond swapping (Fig. 9b).

It is striking to note that the ANNaMo results are at least as good as those obtained with oxDNA, but at a fraction of the computational cost: with the new model, the average (single-CPU-core) walltime required by unbiased MD simulations to observe displacement in systems with toeholds longer than one bead is smaller than one minute at a strand concentration of ≈ 1 mM. However, for shorter toeholds even with ANNaMo TMSD processes become hard to probe due to the exponential dependence of the rate and, unless bond-swapping is enabled, FFS or equivalent techniques are needed. However, as shown in Fig. 9b, if bond-swapping is enabled, the time-dependence of the displacement rate is weaker, and displacements in one-bead-long toehold are only marginally slower than in the case of longer toeholds. Such a high sampling efficiency makes it possible to investigate, and possibly optimise the kinetics of, more complicated systems featuring many different strand displacement gates[80].

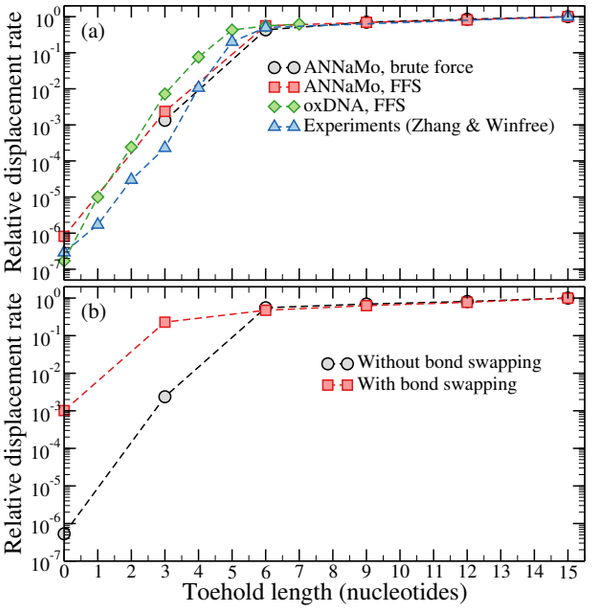


Figure 9. (a) Displacement rates as a function of toehold length (in number of nucleotides) as computed in simulations (*via* brute-force or FFS calculations) and experiments[75], relative to the longest-toehold case (7 nucleotides for the oxDNA data[76], 15 nucleotides in all other cases). (b) The same quantity evaluated with ANNaMo, with and without bond swapping. Errors are always smaller than symbol size.

IV. DISCUSSION AND CONCLUSIONS

This study presents a novel coarse-grained model aimed at simulating folding processes of DNA and RNA nanostructures, ANNaMo. By representing $n = 3$ nucleotides with a single patchy particle, we have achieved a balance between computational efficiency and the ability to capture the interactions that govern folding dynamics, as demonstrated by simulations of DNA hairpins, an RNA pseudoknot, and an RNA tile. The model is parametrized using well-established nearest-neighbor models, and can offer insights into the stability and thermodynamics of nucleic acid structures, while opening avenues for the exploration of larger systems and longer timescales.

We showed that the thermodynamic performance of the model is comparable with those of the best nucleotide-level coarse-grained models (oxDNA and oxRNA), while being two orders of magnitude faster. From the kinetic point of view, ANNaMo reproduces the dependence of the rate of toehold-mediated strand displacement processes on toehold length as observed with experiments or nucleotide-level simulations.

While the model focuses on thermodynamics, some structural properties of nucleic acids, such as the different persistence length between single- and double-stranded molecules, are also retained, even if we observe a somewhat large tendency of ANNaMo dsDNA to kink in simple cases like hairpin structures, where the system lacks sufficient constraints to reduce the likelihood of kinking. Finally, we showed that with ANNaMo it is possible to straightforwardly obtain melting curves of larger structures, such as an RNA tile, *via* unbiased simulations. However, in this case we have observed a likely overestimation of the stability of some specific motifs, such as kissing loops.

Possible future applications of the model are the exploration of folding pathways, optimization of single-stranded motifs and origami designs[5,6], vaccine design[81,82], and viral RNA folding and packaging[83,84]. Although the possibility of simulating multi-stranded systems makes it possible to simulate the formation of complex nanostructures such as DNA origami with the aim of understanding and optimising their folding pathways, doing so may require adding additional (coaxial) stacking interactions between the domains[85,86].

ACKNOWLEDGMENTS

We thank Francesco Sciortino for the fruitful discussions. PŠ acknowledges support by the National Science Foundation under Grant DMR-2239518.

DATA AVAILABILITY

The code used to run all simulations, together with selected example input files can be found in Ref.[59]. The data that support the findings of this study are available from the corresponding author upon request.

Appendix A: The functional forms of the interaction potential

The potentials between **topologically bonded beads** (beads linked through the backbone) are:

- The Kremer-Grest force field[53] V_{KG}, which is the sum of a WCA potential and a FENE potential (see Figure 10). This spring-like potential guarantees excluded volume (WCA component) and attraction (FENE component) that mimic the covalent bonds along the strand. In particular, defining r as the distance between bonded beads, the WCA potential is

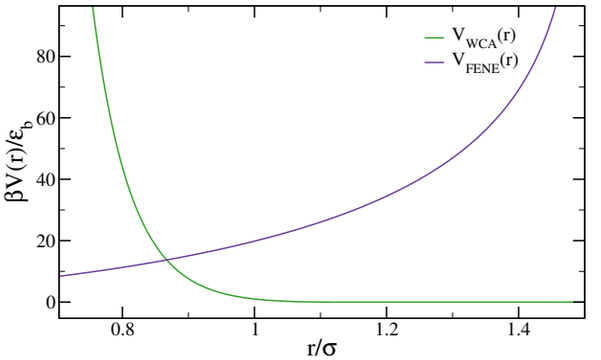


Figure 10. The different potentials acting between bonded beads.

$$V_{WCA}(r) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \varepsilon, & r \le 2^{1/6}\sigma \\ 0, & r > 2^{1/6}\sigma \end{cases} \quad \text{(A1)}$$

where ε and σ are the units of energy and length respectively (both set to 1 in our simulations), while the FENE potential is

$$V_{FENE}(r) = -\frac{1}{2}Kd_0^2 \ln\left[1 - \left(\frac{r}{d_0}\right)^2\right], \quad \text{(A2)}$$

where $d_0 = 1.5\sigma$ and $K = 30\varepsilon/\sigma^2$.

- A three-body potential that tends to align three consecutive beads (see Figure 11):

$$V_{\text{semiflex}}(\theta) = -k_s\left[e^{-\left(\frac{1-\cos\theta}{\xi}\right)^2} - 1\right]. \quad \text{(A3)}$$

where θ is the angle defined by a triplet of bonded particles.

- A term V_{stack} that models the stacking of base in DNA/RNA, acting on the $\vec{a}_1$ directions of consecutive beads:

$$V_{\text{stack}}(\vec{a}_1^{(i)}, \vec{a}_1^{(i+1)}) = \eta(1 - \vec{a}_1^{(i)} \cdot \vec{a}_1^{(i+1)}). \quad \text{(A4)}$$

Non-bonded beads interact through the following potentials:

- A WCA potential, Eq. (A1), to model the excluded-volume interaction;
- An attractive potential that models the hybridization of nucleotides, $V_{\text{sticky}} \cdot V_{\text{direct}}$. Each patch has a position that is given by the position of the bead it is attached to plus $\delta_{pb}\vec{a}_1$, where $\delta_{pb} = 0.65\sigma$, and $V_{\text{sticky}}(r_{pp})$ acts between any two patches, Its functional form was proposed by Stillinger and Weber[54] (see Figure 12) and reads

$$V^{ij}_{\text{sticky}}(r_{pp}) = A\varepsilon_{ij}\left[B\left(\frac{\sigma_s}{r_{pp}}\right)^4 - 1\right]e^{\sigma_s/(r_{pp}-r_c)} \quad \text{(A5)}$$

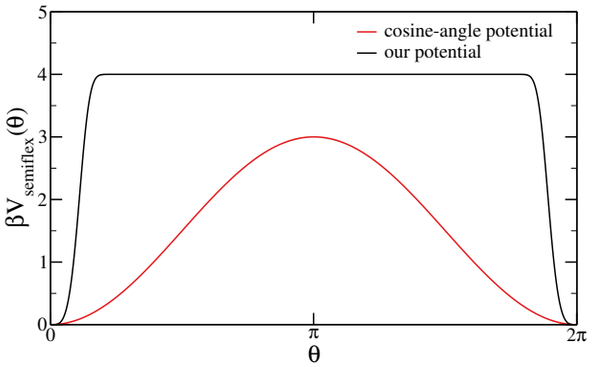


Figure 11. Plot depicting a potential designed to simulate the distinct behaviors of DNA/RNA in single- and double-stranded configurations (black curve). A widely used cosine-angle potential to introduce structural rigidity is included for comparison (red curve). The advantage of our potential is that when there are fewer constraints like in the single-stranded case, the polymer is more likely to rotate freely.

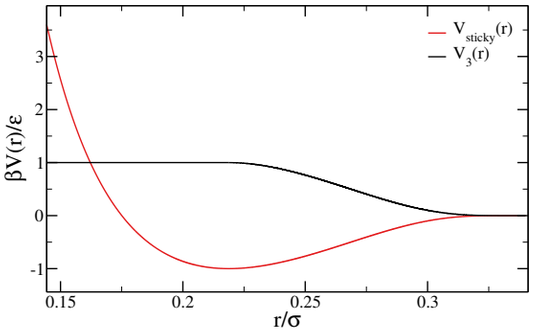


Figure 12. The different potentials acting between non-bonded beads.

where i and j are the bead types and r_{pp} is the distance between the patches and

$$\begin{aligned} B &= \frac{1}{1+4\,(1-r_s)^2}, \\ A &= -\frac{1}{B-1}\frac{1}{e^{1/1-r_s}}, \\ r_c &= \sigma_s r_s. \end{aligned} \tag{A6}$$

The coefficient ε_{ij} modulates the strength of the sticky attraction and depends on what nucleotides are inside the considered beads (see Section II B). We set $\varepsilon_{ij} = 0$ between beads i and j that are first neighbors since loops with lengths shorter than 3 are sterically prohibited[40]. This potential depends essentially on two parameters: σ_s, which defines the minimum of the potential (*i.e.* the radius of the patch) and is set to 0.21875σ, and r_s, which defines the steepness of the potential between the minimum and r_c (set to 0.35), after which $V_{\text{sticky}} = 0$. Thus, two particles are bonded if the relative distance between their patches is less than r_c. The directionality of DNA/RNA is enforced by multiplying V_{sticky} by a term acting on the $\vec{a}_3$ directions of beads:

$$V_{\text{direct}}(\vec{a}_3^{(i)}, \vec{a}_3^{(j)}) = \frac{\vec{a}_3^{(i)} \cdot \vec{a}_3^{(j)}}{2}. \tag{A7}$$

This term makes sure that only antiparallel strands can bind to each other.

- V_{3b}, To ensure the single-bond-per-bead condition, we implement a repulsive three-body interaction V_{3b} which penalizes the formation of triplets of bonded beads[55]. In particular, V_{3b} is designed to almost exactly compensate the gain associated with the formation of a second bond, originating an almost flat energy hypersurface that favors bond swapping even when the bonding energy is much larger than the thermal energy. This repulsive potential is defined as

$$V_{3b} = \lambda \sum_{ijk} \min(\varepsilon_{ij}, \varepsilon_{ik}) V_3(r_{ij}) V_3(r_{ik}) \tag{A8}$$

where the sum runs over all triplets of bonded particles (bead i bonded both with k and j), r_{ij} is the distance between bead i and j, and the minimum between ε_{ij} and ε_{ik} is chosen to favor the removal of the more loosely bonded bead. The value of the parameter λ allows to interpolate between the limits of swapping ($\lambda = 1$) and non-swapping ($\lambda \gg 1$, which here we set to 10) bonds. The pair potential $V_3(r)$ is defined in terms of the normalized $V_{\text{sticky}}(r)$ as

$$V_3(r) = \begin{cases} 1, & r \le \sigma_s \\ -\frac{V^{ij}_{\text{sticky}}(r)}{\varepsilon_{ij}}, & \sigma_s \le r \le r_c \end{cases} \tag{A9}$$

where σ_s is the distance at which $V^{ij}_{sticky}(r)$ has a minimum.

Appendix B: Mapping

Here we show how the free-energy differences prescribed by NN models are mapped onto the strengths of the bead-bead sticky interactions in ANNaMo. In order to do so we adopt the procedure of Ref.[60]: the equilibrium constant associated with the chemical equilibrium of a dimer AB and two beads A and B is given by:

$$K = \frac{[AB]/[\ominus]}{([A]/[\ominus])([B]/[\ominus])} = \frac{\rho_{AB}\rho^{\ominus}}{\rho_A \rho_B} = \exp(-\beta \Delta G^{\ominus}), \tag{B1}$$

where $[\ominus] = 1\,\text{mol dm}^{-3}$ is the standard state concentration, $\rho^{\ominus} = [\ominus]\mathcal{N}_A = 6.022 \times 10^{26}\,\text{m}^{-3}$ is the standard number density and $\Delta G^{\ominus}$ is the standard Gibbs energy for the transformation $A + B \rightleftharpoons AB$ where 50% of the beads have hybridized. Since the functional units of the model are undeformable spheres, they do not possess any internal degree of freedom

and therefore their partition functions can be set equal to 1, and therefore the right-hand side of eq.B1 can be written as

$$q_{AB}\rho^{\circ} = \exp\left(-\beta\Delta G^{\circ}\right). \tag{B2}$$

where q_{AB} is the internal partition function of AB. Knowing that:

$$q_{AB} = 4\pi \int r^2 dr \int d\{\vec{a}\} \int d\{\vec{b}\} e^{-\beta V_{\text{tot}}} \tag{B3}$$

where V_{tot} is the total interaction potential between the two beads, we evaluate the right-hand side of Eq. B3 through a Monte Carlo integration, finding that, for $\varepsilon \gtrsim 3$, $q_{AB} \simeq V_b e^{\alpha\beta\varepsilon}$, where $V_b = 0.0019\,\sigma^3$ and $\alpha = 0.89$ are fitting parameters. Substituting this relation in eq.B2 we find

$$\tilde{V}_b e^{\alpha\beta\varepsilon_{ij}} \sigma^3 \rho^{\circ} = e^{-\beta\Delta G^{\circ}}, \tag{B4}$$

so that

$$\ln\left(\rho^{\circ}\sigma^3\tilde{V}_b\right) + \alpha\beta\varepsilon_{ij} = -\beta\Delta G^{\circ}$$

$$\beta\varepsilon_{ij} = -\frac{1}{\alpha}\left[\beta\Delta G^{\circ} + \ln\left(\rho^{\circ}\sigma^3\tilde{V}_b\right)\right] \tag{B5}$$

Since $\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ}$, we can rewrite the previous equation as:

$$\beta\varepsilon_{ij} = -\frac{1}{\alpha}\left[\frac{1}{k_B T}(\Delta H^{\circ} - T\Delta S^{\circ}) + \ln\left(\rho^{\circ}\sigma^3\tilde{V}_b\right)\right]. \tag{B6}$$

Finally, using eq.2, we obtain:

$$\beta\varepsilon_{ij} = -\frac{1}{\alpha}\left[\left(\frac{\Delta H^{\circ}}{k_B T_{37}}\frac{T_{37}}{T} - \frac{\Delta S^{\circ}}{k_B}\right) + \ln\left(\rho^{\circ}\sigma^3\tilde{V}_b\right)\right]. \tag{B7}$$

Appendix C: Duplex sequences

The sequences used for studying the melting temperatures of the duplexes III D are listed in Table II. We reported just one of the strands forming the duplexes since the other one is just its complementary strand. For each set of sequences, it is specified the size of the simulation box in internal units.

Appendix D: FENE distance

We report the study of the study of the FENE distances for both ds-DNA and ss-DNA. As shown in Figure 13, our model does not distinguish between ss- and ds-chain, both characterized by the same FENE distance. We used this result to convert the persistence length from number of beads to nm: $l_p(\text{nm}) = 0.97\sigma l_p(\#_{\text{beads}})$.

Appendix E: Kink evaluation

Figure 14 reports the cosine angle distribution between adjacent beads found in the study of the persistence length of the ds-DNA. As mentioned in the main text, we define a kink when $\cos\theta_{ij}$ is such that $V_{\text{semiflex}}(\theta_{ij}) - k_s > -k_B T_i$.

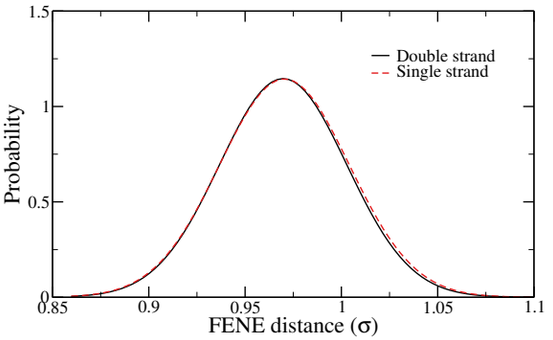


Figure 13. The average FENE distance is $(0.97 \pm 0.03)\sigma$ for both ssDNA and dsDNA.

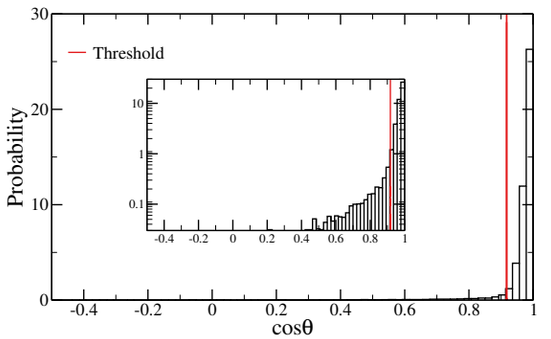


Figure 14. Cosine angle distribution between adjacent beads for ds-DNA. The red line is the threshold below which we consider a kink to have formed. The inset shows the same data on a semi-log scale.

REFERENCES

[1] S. Pitchiaya and Y. Krishnan, "First blueprint, now bricks: Dna as construction material on the nanoscale," Chem. Soc. Rev. **35**, 1111–1121 (2006).
[2] S. Dey, C. Fan, K. V. Gothelf, J. Li, C. Lin, L. Liu, N. Liu, M. A. D. Nijenhuis, B. Saccà, F. C. Simmel, H. Yan, and P. Zhan, "Dna origami," Nature Reviews Methods Primers **1**, 13 (2021).
[3] P. W. K. Rothemund, "Folding dna to create nanoscale shapes and patterns," Nature **440**, 297–302 (2006).
[4] Y. Ke, L. L. Ong, W. M. Shih, and P. Yin, "Three-dimensional structures self-assembled from dna bricks," science **338**, 1177–1183 (2012).
[5] X. Qi, F. Zhang, Z. Su, S. Jiang, D. Han, B. Ding, Y. Liu, W. Chiu, P. Yin, and H. Yan, "Programming molecular topologies from single-stranded nucleic acids," Nature Communications **9**, 4579 (2018).
[6] D. Han, X. Qi, C. Myhrvold, B. Wang, M. Dai, S. Jiang, M. Bates, Y. Liu, B. An, F. Zhang, H. Yan, and P. Yin, "Single-stranded dna and rna origami," Science **358**, eaao2648 (2017).
[7] C. Geary, G. Grossi, E. K. McRae, P. W. Rothemund, and E. S. Andersen, "Rna origami design tools enable cotranscriptional folding of kilobase-sized nanoscaffolds," Nature chemistry **13**, 549–558 (2021).
[8] K. E. Watters, E. J. Strobel, A. M. Yu, J. T. Lis, and J. B. Lucks, "Cotranscriptional folding of a riboswitch at nucleotide resolution," Nature structural & molecular biology **23**, 1124–1131 (2016).
[9] H. Zhang, L. Zhang, A. Lin, C. Xu, Z. Li, K. Liu, B. Liu, X. Ma, F. Zhao, H. Jiang, *et al.*, "Algorithm for optimized mrna design improves stability and immunogenicity," Nature , 1–3 (2023).
[10] C. Du, P. Hu, and L. Ren, "Nucleic acid-based scaffold systems and application in enzyme cascade catalysis," Applied Microbiology and Biotechnology **107**, 9–23 (2023).
[11] Y. Benenson, "Rna-based computation in live cells," Current Opinion in Biotechnology **20**, 471–478 (2009), protein technologies / Systems and syn-

2 beads ($L_{box} = 8\sigma$)	SantaLucia	ANNaMo	Δ	3 beads ($L_{box} = 8\sigma$)	SantaLucia	ANNaMo	Δ
TTTTCG	44	41	-3	CCTTGTTGT	66	66	0
GCAATA	42	42	0	TCTTGGCCT	75	75	0
CGGGCT	68	64	-4	CTTCTTTGG	62	62	0
GGACAG	52	48	-4	CTGTGCACA	72	73	1
AGTAGG	45	43	-2	ACTATTAGG	56	59	3
ATTCTG	40	37	-3	ACCATTCTG	65	66	1
TAGTAC	36	37	1	GGATTAGTA	56	59	3
GCACGT	62	58	-4	ACCGGCTCC	82	82	0
CATCGG	54	51	-3	GTCGGTTGT	72	72	0
GAGAAA	40	36	-4	AACCGCCCC	83	85	2
4 beads ($L_{box} = 8\sigma$)	SantaLucia	ANNaMo	Δ	5 beads ($L_{box} = 10\sigma$)	SantaLucia	ANNaMo	Δ
GGTTGTTCGTAT	73	77	4	CTGGACCGGTGAGTG	85	91	6
TAGCGACATAGA	73	79	6	TCCGCCTTCAATTCC	83	82	-1
ACGGCGCTATGG	85	87	2	GCGACTAAAGGGACT	81	80	-1
GTGCAGGTCGTC	82	85	3	TGGATGGACGCCGAC	88	92	4
ATGAGAGGATAC	70	75	5	ATGTCGTGCAATGTG	80	83	3
ACGATATCCCGG	79	75	-4	GATGATTTAACGCTA	72	77	5
GCCTTTACGCTT	78	77	-1	GAGGCTGTCCCCTCC	88	94	6
TAGAATGGTTAG	66	72	6	TGGCAACTTCGAAAT	79	82	3
CCTCAGTGAAAC	74	76	2	GAGTTAAGCATGCCA	79	76	-3
AGACTGCCTACG	79	84	5	CATCTGCAGTCGGGA	85	88	3
6 beads ($L_{box} = 10\sigma$)	SantaLucia	ANNaMo	Δ	7 beads ($L_{box} = 16\sigma$)	SantaLucia	ANNaMo	Δ
TCTGGAGACCGCCATGAC	88	93	5	GCCGGAGGGTTTTAGTCATTC	84	85	1
AATGAGCCCGCCGCTTAA	90	95	5	GAGTTTGAGCGCCTCATAAAT	81	82	1
TCGTTAAAGATACTGGTT	77	82	5	GCTACCTTCCCGGTCATCCCA	89	92	3
TGCGGGAGACCTAGCCGA	93	94	1	AACCTTAAGCAGGTGAGTGAT	82	91	9
ACCTACGAAATGATGAAC	78	86	8	AACCATTCATGGTTGCCGGAG	86	89	3
TACGCGAAAGGAAAGTGT	83	82	-1	GCCCTGATGCCGAGCAAGCCT	93	93	0
ATCCAGCATGATTTCTCC	81	83	2	AGACTGAGTATCGTCATACCA	80	84	4
TGTATCGAGCTATTACAG	76	80	4	TAGTTGGTCTGCACATTTTGT	81	86	5
GAGGTAGTAGATCCGGTT	81	87	6	ATTAACAGACATGATTTCTTA	73	80	7
TTACTGAAACGGATCCAG	80	83	3	GCCTGCTACTGGTCCGCGACC	92	95	3

8 beads ($L_{box} = 16\sigma$)	SantaLucia	ANNaMo	Δ
GCGCGTCCCCGGTAACGGCGCAGC	99	104	5
CGTCTAGGGGGGTCTGAACCCTCC	91	91	0
AGTTACCTTACCGCTAAACTTGTT	82	87	5
TAGCACGTGCGGCACAACGGTCCA	94	98	4
TGGAGATAGCACCGCTAAGTCATA	84	91	7
CGCACGCCCGTGCACGCCCGGAAG	99	106	7
TCCTGTTTTGAAAGTCTGGCTCGT	86	87	1
GTCCGACAATTTTAGACGATTGGT	82	83	1
CATATCAGAGTTGGAGGAGGCTGC	86	91	5
GTGGCTGCTACGCGTAGGCCTAGG	92	101	9

Table II. Sequences used to simulate duplexes. We have only reported one of the strands since the other is its complement. L_{box} represents the size of the simulation box

thetic biology.

[12] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," Journal of the American Chemical Society **117**, 5179–5197 (1995), https://doi.org/10.1021/ja00124a002.

[13] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "Charmm: A program for macromolecular energy, minimization, and dynamics calculations," Journal of Computational Chemistry **4**, 187–217 (1983), https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.540040211.

[14] 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," Chemical Reviews **118**, 4177–4338 (2018).

[15] M. Zgarbová, J. Sponer, M. Otyepka, T. E. Cheatham III, R. Galindo-Murillo, and P. Jurecka, "Refinement of the sugar–phosphate backbone torsion beta for amber force fields improves the description of z-and b-dna," Journal of chemical theory and computation **11**, 5723–5736 (2015).

[16] M. Parisien and F. Major, "The mc-fold and mc-sym pipeline infers rna structure from sequence data," Nature **452**, 51–55 (2008).

[17] A. Savelyev and G. A. Papoian, "Molecular renormalization group coarse-graining of polymer chains: Application to double-stranded dna," Biophysical Journal **96**, 4044–4052 (2009).

[18] M. Paliy, R. Melnik, and B. A. Shapiro, "Coarse-graining rna nanostructures for molecular dynamics simulations," Physical Biology **7**, 036001 (2010).

[19] S. Pasquali and P. Derreumaux, "Hire-rna: A high resolution coarse-grained energy model for rna," The Journal of Physical Chemistry B **114**, 11957–11966 (2010).

[20]T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, "Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model," The Journal of Chemical Physics **134**, 085101 (2011), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.3552946/14904792/085101_1_online.pdf.

[21]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," The Journal of Chemical Physics **137**, 135101 (2012), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4754132/15454768/135101_1_online.pdf.

[22]D. M. Hinckley, G. S. Freeman, J. K. Whitmer, and J. J. de Pablo, "An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization," The Journal of Chemical Physics **139**, 144903 (2013), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4822042/14787272/144903_1_online.pdf.

[23]T. Cragnolini, P. Derreumaux, and S. Pasquali, "Coarse-grained simulations of rna and dna duplexes," The Journal of Physical Chemistry B **117**, 8047–8060 (2013).

[24]Z. Xia, D. R. Bell, Y. Shi, and P. Ren, "Rna 3d structure prediction by using a coarse-grained model and experimental data," The Journal of Physical Chemistry B **117**, 3135–3144 (2013).

[25]N. A. Denesyuk and D. Thirumalai, "Coarse-grained model for predicting rna folding thermodynamics," The Journal of Physical Chemistry B **117**, 4901–4911 (2013).

[26]P. Šulc, F. Romano, T. E. Ouldridge, J. P. K. Doye, and A. A. Louis, "A nucleotide-level coarse-grained model of RNA," The Journal of Chemical Physics **140**, 235102 (2014), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4881424/15479634/235102_1_online.pdf.

[27]N. Korolev, D. Luo, A. P. Lyubartsev, and L. Nordenskiöld, "A coarse-grained dna model parameterized from atomistic simulations by inverse monte carlo," Polymers **6**, 1655–1675 (2014).

[28]M. Maciejczyk, A. Spasic, A. Liwo, and H. A. Scheraga, "Dna duplex formation with a coarse-grained model," Journal of Chemical Theory and Computation **10**, 5020–5035 (2014).

[29]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," Journal of Chemical Theory and Computation **10**, 2891–2896 (2014).

[30]L. Rovigatti, P. Šulc, I. Z. Reguly, and F. Romano, "A comparison between parallelization approaches in molecular dynamics simulations on gpus," Journal of Computational Chemistry **36**, 1–8 (2015), https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.23763.

[31]M. R. Machado and S. Pantano, "Exploring laci–dna dynamics by multiscale simulations using the sirah force field," Journal of Chemical Theory and Computation **11**, 5012–5023 (2015).

[32]J. J. Uusitalo, H. I. Ingólfsson, P. Akhshi, D. P. Tieleman, and S. J. Marrink, "Martini coarse-grained force field: Extension to dna," Journal of Chemical Theory and Computation **11**, 3932–3945 (2015).

[33]P. D. Dans, J. Walther, H. Gómez, and M. Orozco, "Multiscale simulation of dna," Current Opinion in Structural Biology **37**, 29–45 (2016), theory and simulation • Macromolcular machines.

[34]I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case, and M. Orozco, "Parmbsc1: a refined force field for dna simulations," Nature Methods **13**, 55–58 (2016).

[35]D. Chakraborty, N. Hori, and D. Thirumalai, "Sequence-dependent three interaction site model for single- and double-stranded dna," Journal of Chemical Theory and Computation **14**, 3763–3779 (2018).

[36]C. Maffeo and A. Aksimentiev, "MrDNA: a multi-resolution model for predicting the structure and dynamics of DNA systems," Nucleic Acids Research **48**, 5135–5146 (2020), https://academic.oup.com/nar/article-pdf/48/9/5135/33220929/gkaa200.pdf.

[37]M. DeLuca, T. Ye, M. Poirier, Y. Ke, C. Castro, and G. Arya, "Mechanism of dna origami folding elucidated by mesoscopic simulations," bioRxiv (2023), 10.1101/2023.06.20.545758.

[38]A. Sengar, T. E. Ouldridge, O. Henrich, L. Rovigatti, and P. Šulc, "A primer on the oxdna model of dna: When to use it, how to simulate it and how to interpret the results," Frontiers in Molecular Biosciences **8** (2021), 10.3389/fmolb.2021.693710.

[39]D. Poland and H. A. Scheraga, "Occurrence of a phase transition in nucleic acid models," The Journal of Chemical Physics **45**, 1464–1469 (2004).

[40]J. SantaLucia and D. Hicks, "The thermodynamics of dna structural motifs," Annual Review of Biophysics and Biomolecular Structure **33**, 415–440 (2004), pMID: 15139820, https://doi.org/10.1146/annurev.biophys.32.110601.141800.

[41]H. T. Allawi and J. SantaLucia, "Nearest-neighbor thermodynamics of internal a·c mismatches in dna: sequence dependence and ph effects," Biochemistry **37**, 9435–9444 (1998).

[42]H. Allawi and J. Santalucia, "Thermodynamics of internal c·t mismatches in dna," Nucleic acids research **26**, 2694–701 (1998).

[43]N. Peyret, P. Seneviratne, H. Allawi, and J. Santalucia, "Nearest-neighbor thermodynamics and nmr of dna sequences with internal a.a, c.c, g.g, and t.t mismatches," Biochemistry **38**, 3468–77 (1999).

[44]H. Allawi and J. Santalucia, "Thermodynamics and nmr of internal gt mismatches in dna," Biochemistry **36**, 10581–94 (1997).

[45]N. Peyret, P. Seneviratne, H. Allawi, and J. Santalucia, "Nearest-neighbor thermodynamics and nmr of dna sequences with internal a.a, c.c, g.g, and t.t mismatches," Biochemistry **38**, 3468–77 (1999).

[46]D. H. Turner and D. H. Mathews, "NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure," Nucleic Acids Research **38**, D280–D282 (2009), https://academic.oup.com/nar/article-pdf/38/suppl_1/D280/11217894/gkp892.pdf.

[47]J. M. Huguet, C. V. Bizarro, N. Forns, S. B. Smith, C. Bustamante, and F. Ritort, "Single-molecule derivation of salt dependent base-pair free energies in dna," Proceedings of the National Academy of Sciences **107**, 15431–15436 (2010), https://www.pnas.org/doi/pdf/10.1073/pnas.1001454107.

[48]J. H. Bae, J. Z. Fang, and D. Y. Zhang, "High-throughput methods for measuring DNA thermodynamics," Nucleic Acids Research **48**, e89–e89 (2020), https://academic.oup.com/nar/article-pdf/48/15/e89/33697352/gkaa521.pdf.

[49]J. Zuber, S. J. Schroeder, H. Sun, D. H. Turner, and D. Mathews, "Nearest neighbor rules for RNA helix folding thermodynamics: improved end effects," Nucleic Acids Research **50**, 5251–5262 (2022), https://academic.oup.com/nar/article-pdf/50/9/5251/43774084/gkac261.pdf.

[50]B. E. Snodin, F. Randisi, M. Mosayebi, P. Šulc, J. S. Schreck, F. Romano, T. E. Ouldridge, R. Tsukanov, E. Nir, A. A. Louis, *et al.*, "Introducing improved structural properties and salt dependence into a coarse-grained model of dna," J. Chem. Phys. **142** (2015).

[51]M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, "Stretching dna with optical tweezers," Biophysical Journal **72**, 1335–1346 (1997).

[52]E. Roth, A. Glick Azaria, O. Girshevitz, A. Bitler, and Y. Garini, "Measuring the conformation and persistence length of single-stranded dna using a dna origami structure," Nano Letters **18**, 6703–6709 (2018).

[53]G. S. Grest and K. Kremer, "Molecular dynamics simulation for polymers in the presence of a heat bath," Phys. Rev. A **33**, 3628–3631 (1986).

[54]T. A. Weber and F. H. Stillinger, "Local order and structural transitions in amorphous metal-metalloid alloys," Phys. Rev. B **31**, 1954–1963 (1985).

[55]F. Sciortino, "Three-body potential for simulating bond swaps in molecular dynamics," The European Physical Journal E **40**, 3 (2017).

[56]M. T. Record, C. F. Anderson, and T. M. Lohman, "Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity," Quarterly Reviews of Biophysics **11**, 103–178 (1978).

[57]J. Russo, P. Tartaglia, and F. Sciortino, "Reversible gels of patchy particles: Role of the valence," The Journal of Chemical Physics **131**, 014504 (2009).

[58]E. Poppleton, M. Matthies, D. Mandal, F. Romano, P. Šulc, and L. Rovigatti, "oxdna: coarse-grained simulations of nucleic acids made simple," Journal of Open Source Software **8**, 4693 (2023).

[59]F. Tosti Guerra and L. Rovigatti, "lorenzo-rovigatti/annamo_files: v1.0," (2024).

[60]A. Reinhardt and D. Frenkel, "Dna brick self-assembly with an off-lattice potential," Soft Matter **12**, 6253–6260 (2016).

[61]A. Peselis and A. Serganov, "Structure and function of pseudoknots involved in gene expression control," WIREs RNA **5**, 803–822 (2014).

[62]K. Neupane, M. Zhao, A. Lyons, S. Munshi, S. M. Ileperuma, D. B. Ritchie, N. Q. Hoffer, A. Narayan, and M. T. Woodside, "Structural dynamics of single sars-cov-2 pseudoknot molecules reveal topologically distinct conformers," Nature Communications **12**, 4749 (2021).

[63]M. Kucharík, I. L. Hofacker, P. F. Stadler, and J. Qin, "Pseudoknots in RNA folding landscapes," Bioinformatics **32**, 187–194 (2015), https://academic.oup.com/bioinformatics/article-pdf/32/2/187/49016517/bioinformatics_32_2_187.pdf.

[64]S. R. Eddy, "How do rna folding algorithms work?" Nature Biotechnology **22**, 1457–1458 (2004).

[65]D. P. GIEDROC and C. A. THEIMER, "Contribution of the intercalated adenosine at the helical junction to the stability of the gag-pro frameshifting pseudoknot from mouse mammary tumor virus," RNA **6**, 409–421 (2000).

[66]E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, "Design and self-assembly of two-dimensional dna crystals," Nature **394**, 539–544 (1998).

[67]E. Poppleton, J. Bohlin, M. Matthies, S. Sharma, F. Zhang, and P. Šulc, "Design, optimization and analysis of large dna and rna nanostructures through interactive visualization, editing and molecular simulation," Nucleic Acids Research **48**, e72–e72 (2020).

[68]J. C. Paillart, E. Skripkin, B. Ehresmann, C. Ehresmann, and R. Marquet, "A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic hiv-1 rna." Proceedings of the National Academy of Sciences **93**, 5572–5577 (1996).

[69]E. Bindewald, R. Hayes, Y. G. Yingling, W. Kasprzak, and B. A. Shapiro, "Rnajunction: a database of rna junctions and kissing loops for three-dimensional structural analysis and nanodesign," Nucleic Acids Research **36**, D392–D397 (2008).

[70]D. Liu, C. W. Geary, G. Chen, Y. Shao, M. Li, C. Mao, E. S. Andersen, J. A. Piccirilli, P. W. K. Rothemund, and Y. Weizmann, "Branched kissing loops for the construction of diverse rna homooligomeric nanostructures," Nature Chemistry **12**, 249–259 (2020).

[71]T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, "Extracting bulk properties of self-assembling systems from small simulations," Journal of Physics: Condensed Matter **22**, 104102 (2010).

[72]F. C. Simmel, B. Yurke, and H. R. Singh, "Principles and applications of nucleic acid strand displacement reactions," Chemical reviews **119**, 6326–6369 (2019).

[73]T. Luo, S. Fan, Y. Liu, and J. Song, "Information processing based on dna toehold-mediated strand displacement (tmsd) reaction," Nanoscale **13**, 2100–2112 (2021).

[74]N. Wang, Y. Jiang, K. Nie, D. Li, H. Liu, J. Wang, C. Huang, and C. Li, "Toehold-mediated strand displacement reaction-propelled cascade dnazyme amplifier for microrna let-7a detection," Chinese Chemical Letters **34**, 107906 (2023).

[75]D. Y. Zhang and E. Winfree, "Control of dna strand displacement kinetics using toehold exchange," Journal of the American Chemical Society **131**, 17303–17314 (2009).

[76]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 Research **41**, 10641–10658 (2013).

[77]A. Walbrun, T. Wang, M. Matthies, P. Šulc, F. C. Simmel, and M. Rief, "Single-molecule force spectroscopy of toehold-mediated strand displacement," bioRxiv (2024), 10.1101/2024.01.16.575816, https://www.biorxiv.org/content/early/2024/01/17/2024.01.16.575816.full.pdf.

[78]R. J. Allen, C. Valeriani, and P. R. Ten Wolde, "Forward flux sampling for rare event simulations," Journal of physics: Condensed matter **21**, 463102 (2009).

[79]S. Hussain and A. Haji-Akbari, "Studying rare events using forward-flux sampling: Recent breakthroughs and future outlook," The Journal of Chemical Physics **152**, 060901 (2020), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.5127780/13928672/060901_1_online.pdf.

[80]L. Qian, E. Winfree, and J. Bruck, "Neural network computation with dna strand displacement cascades," nature **475**, 368–372 (2011).

[81]K. Leppek, G. W. Byeon, W. Kladwang, H. K. Wayment-Steele, C. H. Kerr, A. F. Xu, D. S. Kim, V. V. Topkar, C. Choe, D. Rothschild, G. C. Tiu, R. Wellington-Oguri, K. Fujii, E. Sharma, A. M. Watkins, J. J. Nicol, J. Romano, B. Tunguz, F. Diaz, H. Cai, P. Guo, J. Wu, F. Meng, S. Shi, E. Participants, P. R. Dormitzer, A. Solórzano, M. Barna, and R. Das, "Combinatorial optimization of mrna structure, stability, and translation for rna-based therapeutics," Nature Communications **13**, 1536 (2022).

[82]Y.-A. Kim, K. Mousavi, A. Yazdi, M. Zwierzyna, M. Cardinali, D. Fox, T. Peel, J. Coller, K. Aggarwal, and G. Maruggi, "Computational design of mrna vaccines," Vaccine (2023), https://doi.org/10.1016/j.vaccine.2023.07.024.

[83]A. L. Božič, C. Micheletti, R. Podgornik, and L. Tubiana, "Compactness of viral genomes: effect of disperse and localized random mutations," Journal of Physics: Condensed Matter **30**, 084006 (2018).

[84]D. Vaupotič, A. Rosa, L. Tubiana, and A. Božič, "Scaling properties of RNA as a randomly branching polymer," The Journal of Chemical Physics **158**, 234901 (2023).

[85]S. Gambietz, L. J. Stenke, and B. Saccà, "Sequence-dependent folding of monolayered dna origami domains," Nanoscale **15**, 13120–13132 (2023).

[86]A. Cumberworth, D. Frenkel, and A. Reinhardt, "Simulations of dna-origami self-assembly reveal design-dependent nucleation barriers," Nano Letters **22**, 6916–6922 (2022), pMID: 36037484.