Performing SELEX experiments in silico

J. A. J. Wondergem, H. Schiessel, and M. Tompitak

Citation: The Journal of Chemical Physics **147**, 174101 (2017);
View online: https://doi.org/10.1063/1.5001394
View Table of Contents: http://aip.scitation.org/toc/jcp/147/17
Published by the American Institute of Physics

Performing SELEX experiments *in silico*

J. A. J. Wondergem, H. Schiessel, and M. Tompitak[a)]
Institute Lorentz for Theoretical Physics, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands

(Received 24 August 2017; accepted 13 October 2017; published online 1 November 2017)

Due to the sequence-dependent nature of the elasticity of DNA, many protein-DNA complexes and other systems in which DNA molecules must be deformed have preferences for the type of DNA sequence they interact with. SELEX (Systematic Evolution of Ligands by EXponential enrichment) experiments and similar sequence selection experiments have been used extensively to examine the (indirect readout) sequence preferences of, e.g., nucleosomes (protein spools around which DNA is wound for compactification) and DNA rings. We show how recently developed computational and theoretical tools can be used to emulate such experiments *in silico*. Opening up this possibility comes with several benefits. First, it allows us a better understanding of our models and systems, specifically about the roles played by the simulation temperature and the selection pressure on the sequences. Second, it allows us to compare the predictions made by the model of choice with experimental results. We find agreement on important features between predictions of the rigid base-pair model and experimental results for DNA rings and interesting differences that point out open questions in the field. Finally, our simulations allow application of the SELEX methodology to systems that are experimentally difficult to realize because they come with high energetic costs and are therefore unlikely to form spontaneously, such as very short or overwound DNA rings. *Published by AIP Publishing.* https://doi.org/10.1063/1.5001394

I. INTRODUCTION

Over the past 25 years, SELEX (Systematic Evolution of Ligands by EXponential enrichment) experiments have proven a valuable tool in identifying DNA and RNA sequences with high affinity for a large range of target molecules. This affinity can be based on any number of properties of the nucleic acids, such as sequence-specific binding of the target or an RNA's ability to form stem loops. SELEX experiments have found many of their applications in clinical research: to examine the tendency of prospective therapeutic compounds to target specific genomic sequences or designing RNA molecules that themselves interfere with the functioning of certain pathogens. (For a review, see Ref. 1.)

We will focus on the basic mechanics (elasticity and intrinsic shape) of double-stranded DNA molecules and their consequent affinity for certain complexes in which the DNA needs to be deformed. Various DNA-binding proteins are known to have DNA affinities that are dependent on the intrinsic curvature and stiffness of the underlying nucleotide sequence, such as the catabolite activator protein,[2] the TATA-binding factor,[3–5] and other parts of the transcriptional machinery,[6–8] as well as regulatory[9–12] and architectural proteins.[8,13]

However, the archetypical example is the nucleosome, a protein spool around which genomic DNA in eukaryotes is wrapped in order to compactify it.[14] The positioning of these protein spools along a genome influences the packaging of the DNA and thereby the expression of genes, as wrapped-up DNA cannot readily be read out.[15] Since DNA needs to be strongly bent in order to wrap into a nucleosome, the nucleosomal structure has a preference for sequences that facilitate this deformation. This leads to significant effects of the underlying DNA sequence on the positioning and dynamics of nucleosomes.[16]

In this context, SELEX experiments have been used to look for DNA sequences with high affinity to the nucleosome[17–19] (as well as the archaeal "nucleosome"[20]). In similar endeavors, the SELEX method has been used to look for intrinsically curved sequences[21] and to assess the sequence preferences of DNA rings.[22]

In such SELEX experiments, a pool of random DNA molecules is synthesized (either fully randomly or randomly drawn from genomic sequences[23]), and these random molecules are mixed with molecules of the target type, competing to bind to them. The DNA molecules with the highest affinity will be the most likely candidates to bind to the targets. After some time, the DNA-target complexes are extracted from the mixture, leaving behind a fraction of the DNA molecules that have a lower average affinity and keeping a fraction with higher affinity.

By repeating this process in multiple rounds, the selective pressure on the DNA sequences increases and we end up with a smaller and smaller pool of higher and higher affinity sequences. In such a manner, the Widom 601 sequence[19] of high nucleosome affinity was discovered, and the dinucleotide probability distributions of DNA rings were mapped.[22] Although not the same on a technical level, similar experiments have been used to map the sequence preferences of nucleosomes.[24–29] Mapping such preferences is not only an interesting goal in itself but these preferences can also be used to model sequence-dependent nucleosome affinity.

[a)]Electronic mail: tompitak@lorentz.leidenuniv.nl

Such models can in turn be employed to gain insight into the mechanical signals encoded into genomic DNA sequences.[25,27,30,31]

Recently, a computational method has been published that also enables mapping of such sequence preferences.[32] Dubbed Mutation Monte Carlo (MMC), the method utilizes standard Monte Carlo simulations to sample the Boltzmann distribution associated to a modeled DNA system such as the nucleosome and adds as a novel feature Monte Carlo moves that mutate the DNA sequence. Given a suitable model of the system of interest, this technique allows an understanding of the sequence preferences of the system from a theoretical point of view.

The MMC method shares many similarities with the experimental SELEX method. It samples DNA sequences based on their affinity to the target. Doing so at constant finite temperature delivers probability distributions for, e.g., dinucleotides (as in Refs. 30, 32, and 33), and by performing simulated annealing it searches for the sequence with the strongest affinity (Refs. 33 and 34), much as attempted in Ref. 19, leading to the 601 sequence.

However, there is also a major difference between the *in silico* method and the experimental protocols. The MMC simulation is performed at a particular temperature, which determines how stringently it selects for low-energy states and hence for high-affinity sequences. This temperature is necessarily shared by both the configurational moves that simulate the thermal fluctuations of the system and the mutations. In a SELEX experiment, however, the selection pressure is determined by, among other factors, the number of rounds of selection performed, and the strength of selection on the sequences is decoupled from the temperature at which the experiment is performed. Despite the similarities, this means that a MMC simulation cannot be directly taken as an *in silico* SELEX experiment.

Here we bridge this difference, such that we may apply selective pressure *in silico* at will regardless of the simulation temperature. To do so, we must examine in detail the role played by temperature in the MMC method, which we will do in Secs. II and III. Considering MMC simulations of both nucleosomes and DNA rings, we will find in Sec. IV that the importance of the temperature varies from system to system.

With the tools in hand to perform simulated SELEX experiments, we first emulate the experiment performed by Rosanio *et al.* for rings.[22] In Sec. V, we elucidate the fundamental differences between the (out-of-equilibrium) experiment of Rosanio *et al.* and our idealized equilibrium statistics to show that a comparison is useful. After affirming this, we perform the *in silico* selection in Sec. VI, and we find both broad agreement and some striking differences between the theoretical predictions and the experimental results. Finally, in Sec. VII, we apply our SELEX simulations to tight and overwound rings, which would be difficult to treat experimentally due to the lower rate of the formation of such systems.

II. SELEX AND MMC

In a SELEX experiment, DNA molecules compete to bind to target molecules or, in the case of DNA rings, to form closed rings in a limited amount of time.[22] The probability of a molecule with sequence S to be bound to the target instead of another molecule, assuming equilibrium conditions, is proportional to the Boltzmann weight of that molecule's free energy when bound to the target,

$$P(S) = \frac{1}{Z} e^{-\beta F(S)}, \tag{1}$$

where Z is the partition function, i.e., $\sum_S e^{-F(S)/k_B T}$.

A single round in a SELEX experiment is then very similar to a MMC simulation. When we run a MMC simulation, we are sampling system configurations, i.e., combinations of sequences and spatial configurations (S, θ), according to their Boltzmann distribution,

$$P(S, \theta) = \frac{1}{Z} e^{-\beta E(S,\theta)} \delta(f_c(\theta)). \tag{2}$$

The normalization is provided by the partition function Z, obtained by integrating the numerator over all spatial degrees of freedom and summing over all sequences. In this equation, we have added a delta function to encode for the constraints on the system. In a nucleosome, there are constraints on the spatial degrees of freedom that bind the DNA to the histone core. In a DNA ring, the molecule is constrained to form a loop. The exact form of these constraints may be complex and is captured here by a general constraint function f_c.

When we speak of the affinity of a sequence to a nucleosome or a ring, we do not make reference to any particular spatial configuration. Rather we want to take all of them into account; we need the probability of a given sequence to form a nucleosome or ring, considering the probabilities of all the possible spatial configurations the DNA may take. Then what we wish to calculate is the marginal probability distribution of the sequences,

$$P(S) = \int d\theta P(S, \theta) = \frac{1}{Z} \int d\theta e^{-\beta E(S,\theta)} \delta(f_c(\theta)). \tag{3}$$

This integral will not generally be tractable. In the current work, we rely on the rigid base-pair (RBP) model[35] to provide the energy function $E(S, \theta)$. This energy function is quadratic in the degrees of freedom, making the integral above a Gaussian integral under constraints. This may be solvable for very simple constraint functions, but in general we need to resort to numerical methods such as MMC.

Assuming we have a method to evaluate $P(S)$, we can consider the free energy of a given sequence

$$P(S) = \frac{1}{Z} e^{-\beta F(S)} \rightarrow F(S) = -\frac{1}{\beta} \left(\log(P(S)) - \log(Z) \right). \tag{4}$$

The partition function is generally difficult to determine. In what follows, we will neglect its contribution, meaning that we determine the free energy only up to a constant offset. Similarly, we will simply normalize our probability distributions as required and drop overall factors from our equations.

However, besides this caveat, we are determining the same quantities as we would in a SELEX experiment, at least when considering only a single round. In Sec. III, we address simulating SELEX experiments consisting of multiple rounds.

III. AN EFFECTIVE TEMPERATURE FOR MUTATIONS

As noted, the probability of a given sequence to survive a SELEX round depends on its free energy when bound to the target. Assuming a fraction f is kept after a round of SELEX, the survival probability of a sequence S is

$$P_{\text{surv}}(S) = fe^{-\beta F(S)}. \tag{5}$$

For the sequence to survive multiple rounds, assuming selection criteria are constant from one round to the next, we multiply this probability with itself,

$$P_{\text{surv},n}(S) = f^n e^{-n\beta F(S)} = f^n e^{-\beta'_m F(S)}. \tag{6}$$

The fraction f can, in the case of DNA forming nucleosomes or other complexes, be constrained to be smaller than 1 by mixing together a surplus of DNA molecules with the target proteins.

Apart from the scaling with a sequence-independent prefactor, we see that applying n rounds of SELEX is equivalent to introducing an effective temperature, $T \to T' = T/n$. We call this an effective temperature, since it only applies to the selection of the sequences. In what follows, we will therefore distinguish between β_m, the inverse temperature that is applied to sequence selection (i.e., the mutations in our MMC simulation), and β_s, the inverse temperature of the spatial degrees of freedom. In Eq. (6), the actual physical temperature of the system is not altered. We wish to replicate this effect in our MMC simulations.

The free energy in Eq. (4) depends on the simulation temperature and, as noted in the introduction, this temperature governs both the selection of sequences and the selection of spatial configurations during the simulation. However, there is nothing to stop us from tweaking the temperature after marginalizing out the spatial degrees of freedom. If we wish to calculate $P(S)$ at some temperature T' other than the simulation temperature T, we may simply write

$$\begin{aligned} P_{T'_m}(S) &= e^{-\beta'_m F(S)} = \left(e^{-\beta_m F(S)}\right)^{\beta'_m/\beta_m} \\ &= P_{T_m}(S)^{\beta'_m/\beta_m}, \end{aligned} \tag{7}$$

where the temperature subscript to $P(S)$ denotes an *effective* temperature for the mutation moves only. Note that this is distinct from changing the actual simulation temperature, in which case we must write

$$P_{T'}(S) = \int d\theta e^{-\beta' E(S,\theta)} \delta(f_c(\theta)) \tag{8}$$

$$= \int d\theta (e^{-\beta E(S,\theta)})^{\beta'/\beta} \delta(f_c(\theta)). \tag{9}$$

The question of how this expression scales with T' does not have a straightforward answer and depends on the constraints placed upon the system, as we will see.

Assuming we can calculate $P(S)$, Eq. (7) allows us to decouple the selective pressure on the sequences from the simulation temperature, in a manner entirely analogous to how a SELEX experiment introduces an effective temperature for the sequence selection. Furthermore, we are not restricted to temperatures that are integer fractions of the physical temperature; we may choose T' as we like, even a temperature larger than the physical one.

IV. EFFECTIVE TEMPERATURE AND SEQUENCE PREFERENCES

For Eq. (7) to be of use, we need a tractable way to calculate $P(S)$. The MMC method enables us to sample the Boltzmann distribution in sequence space for the system of interest, but sampling the full space of all possible sequences is still an impossible task for systems like the nucleosome, due to the large number of sequences.

The standard way of gaining insight into the sequence preferences of a system is by considering the probability distributions of short subsequences in the full sequence, most commonly those of dinucleotides,[22,25–27,32,33] which is a far more tractable problem. Those distributions capture much of the information about a system's preferences, and they can in fact be employed in calculating the affinity of sequences, if we make some simplifying assumptions. Following Refs. 25 and 30, we assume only short-range correlations in the sequence preferences of our systems, such that we may write

$$P(S) = P(S_1)P(S_2|S_1)\prod_{i=3}^{N} P(S_i|S_{i-1} \cap S_{i-2}), \tag{10}$$

where S_i are the individual nucleotides that make up the DNA sequence. This expression for $P(S)$ (the trinucleotide model from Ref. 30) assumes that the probabilities of the individual nucleotides are only strongly correlated with their nearest and next-nearest neighbours, i.e., the probability of S_i depends only on S_{i-1} and S_{i-2}. This assumption was extensively tested in Ref. 30. Using Eq. (10), we may sample the probability distributions of trinucleotides in our MMC simulation and from there calculate the probability or free energy of an entire sequence.

With this method for calculating $P(S)$ in hand, we can now gather an ensemble of sequences at a different mutation temperature by running a MMC simulation in sequence space only, but where we reject or accept mutations (within the Metropolis-Hastings algorithm) based on the adjusted probabilities given by Eq. (7).

From this new sequence ensemble, we can then once again derive dinucleotide distributions to study. Comparing the distributions found using this method, with the original ones from the single-temperature MMC simulation, we may assess separately the effects of changing the mutation temperature and the spatial temperature.

We modeled DNA using the rigid base-pair model[35] with the standard hybrid parameterization.[36] We ran MMC simulations of nucleosomes (modeled using the Eslami-Mossallam nucleosome model[32]) and rings (modeled by connecting the first and last base pairs of the DNA using the standard sequence-dependent elasticity of the rigid base-pair model) at three different temperatures: room temperature, 1/2 of room temperature, and 1/4 of room temperature. Then we used the method just described to independently alter the mutation temperature. The results are presented in Fig. 1.

The distributions for A/T-rich dinucleotides (a common set to study due to the strong preferences shown by the nucleosome for the positions of these dinucleotides) for the ring and the nucleosome show an interesting difference. In Figs. 1(d)–1(f), we see that the distributions we find for the ring depend

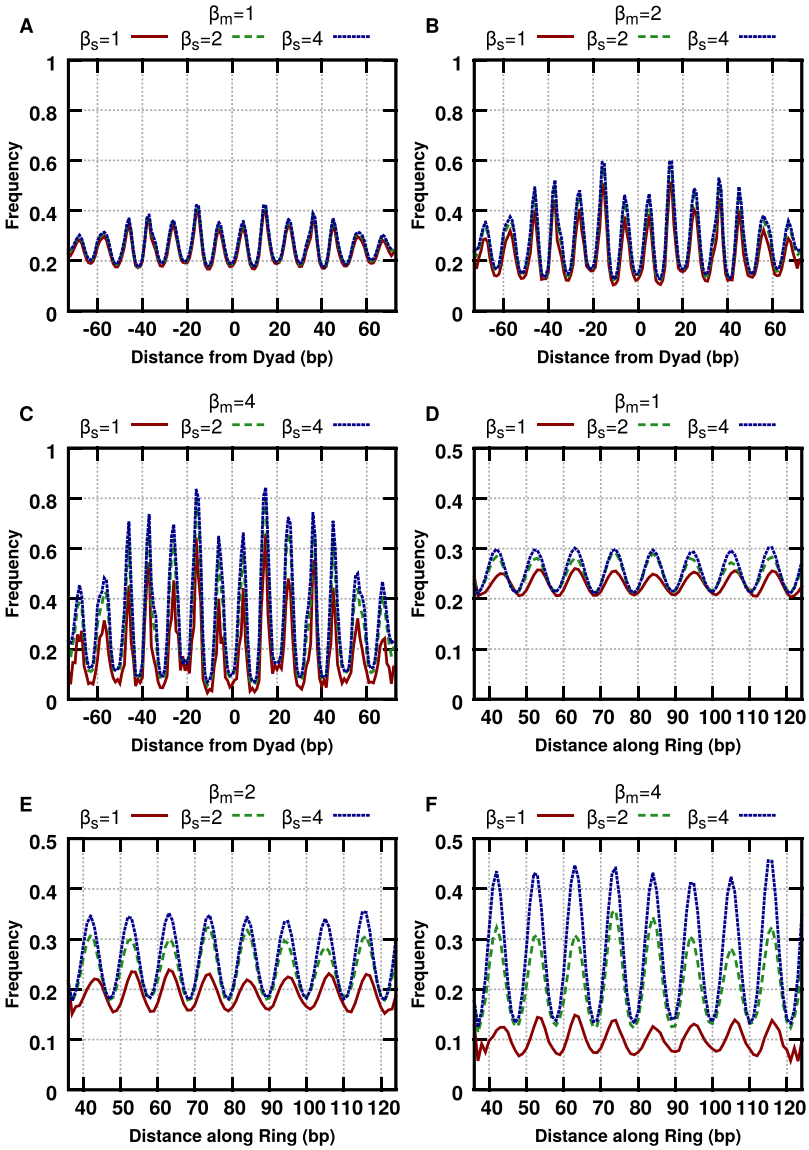

FIG. 1. Distributions for AT-rich dinucleotides (AA, AT, TA, and TT) along the nucleosome [(a)–(c)] and the ring [(d)–(f)] biased by the locking sequence from Rosanio *et al.*,[22] for different combinations of mutation temperature (β_m) and spatial temperature (β_s). The distributions are grouped by mutation temperature in order to illuminate the different effects of spatial temperature on the preferences of the nucleosome and the ring. The effect of multiple rounds of SELEX would be to raise β_m while keeping β_s constant, so one would consider the curves of the same color in successive plots.

strongly not just on the mutation temperature β_m but also on the spatial temperature β_s. For the nucleosome, however, we see in Figs. 1(a)–1(c) a strong dependence on β_m but a far weaker dependence on β_s.

This difference can be understood in terms of the entropic contribution to the free energies of the systems. Considering a given sequence S, its free energy has a contribution from the average internal energy of a system and from the entropy (denoted here by Σ to distinguish it from the sequence S),

$$F(S) = \langle E(S)\rangle - T_s\Sigma(S), \tag{11}$$

where T_s is the spatial temperature, as we are considering the system with a given sequence S.

Since the entropy is a measure of the part of the configuration space that can be accessed with reasonable probability by the system, it in principle depends on the sequence. For example, for a completely free DNA molecule, a stiff sequence will limit the possible spatial configurations of the molecule more than a sequence that bends very easily. Changing the spatial temperature affects the accessible part of state space, and hence the contribution $T_s\Sigma(S)$, in a sequence-dependent manner.

The average energy $\langle E(S)\rangle$ also depends on temperature, but in a straightforward, sequence-independent manner. It represents the internal potential energy plus the thermal energy, simply given by the equipartition theorem,

$$\langle E(S)\rangle = E_0(S) + \frac{N}{2}k_BT_s, \tag{12}$$

where N is the number of degrees of freedom.

The dependence of the sequence preferences of DNA rings we find in Figs. 1(d)–1(f) is thus an entropic effect. At lower temperatures, the ring will be constrained to a smaller set of configurations, but how many depends on what the stiffness of the DNA sequence allows. Hence, lowering the spatial temperature increases the differences in affinity between sequences, leading to the larger amplitudes in Figs. 1(d)–1(f).

For the nucleosome, the effect is much smaller. Apparently, the entropic contribution $T_s\Sigma(S)$ is not strongly sequence-dependent in this case. This was expected: because the nucleosome is a strongly constrained system, the part of

configuration space that the DNA is allowed to sample is determined to a much larger degree by the constraints on the system than by the elastic properties of the DNA itself. This was already anticipated in studies including Refs. 32 and 37, where the entropic contribution to the free energy of the nucleosome was neglected entirely. Using our new methodology, we are able to directly verify that this assumption is justified. However, we must conclude that the assumption does not hold for systems that are not as tightly constrained as the nucleosome, such as, for instance, DNA rings.

V. AN *IN SILICO* SELEX EXPERIMENT FOR RINGS

Having developed the methodology to perform SELEX experiments *in silico*, we would like to compare the results of such computational treatments to experimental results. The most promising experiment to compare with is that of Rosanio *et al.*,[22] the only experiment making use of completely random sequences for which the statistics we are interested in have been reported.

Rosanio *et al.* performed a SELEX experiment in which fragments consisting of 126 base pairs of DNA were made to cyclize into rings. Linear DNA fragments randomly sample bent configurations due to thermal fluctuations, and if the two ends of a fragment meet, a ligation reaction may fuse them together, creating a closed ring. The probability of a given DNA fragment cyclizing depends on its affinity to form a ring: a stiff sequence is less likely to cyclize and survive a selection round than an easily bendable one; the same holds for an intrinsically straight molecule compared with an intrinsically bent one. To gain insight into the sequence preferences of rings, Rosanio *et al.* fixed 36 of the 126 base pairs to contain a predetermined sequence with a known preference for bending in one direction. This biased the direction of ring formation, such that the preferences of a ring bent in a specific direction could be mapped.

We wish to mimic this experiment *in silico* by performing a MMC simulation of a DNA ring, with 36 base pairs fixed to the same sequence used by Rosanio *et al.*, and the rest free to mutate. We found that the RBP model correctly captures the fact that the 36-base-pair locking sequence biases the bending direction in the ring. Figure 2 shows histograms of the rotational states (measuring the rotation of the ring around its own backbone) of DNA rings with locking sequences, and the other 90 base pairs made into homogeneous (sequence-averaged) DNA without a coherent bending preference, sampled during a standard Monte Carlo simulation. We define the rotational state as the signed angle between a vector perpendicular to the ring at an arbitrary point and the plane in which the ring (approximately) lies. The top panel shows the results using the Rosanio sequence; the bottom uses the artificially designed, very strongly intrinsically bent sequence from Ref. 34. The fixed sequences significantly bias the ring to a subrange of rotational states; the artificially designed sequence biases far more strongly than does the Rosanio sequence, for which reason we will employ it later on. As a side remark, note that for the Rosanio sequence, the energy landscape as a function of the rotational angle shows an interesting asymmetry: it is ratchet-shaped. As explained in Refs. 38 and 39, a DNA ring

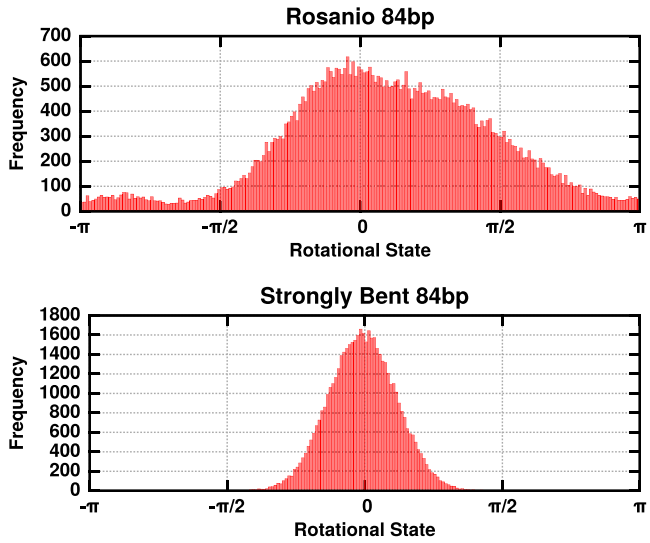

FIG. 2. Histograms of rotational states (around its length axis) of an RBP DNA molecule forced into a ring, sampled during a standard Monte Carlo simulation, for two separate locking sequences consisting of 36 base pairs. The bias introduced using the sequence from Rosanio *et al.*[22] is shown in the top panel. The bottom panel shows the bias produced using an arbitrarily selected 36-base-pair subsequence of the strongly curved 84-base-pair sequence from Ref. 34.

with such a feature can be made to twirl around its backbone via a periodic change in temperature, thus acting as a molecular motor.

Before we present the results of our MMC simulation of the Rosanio *et al.* experiment, it is instructive to first discuss some significant differences.

Ideally, as discussed around Eq. (5), in a SELEX experiment, the survival probability is Boltzmann-distributed. In the cyclization assay, the DNA fragments are initially in rapid equilibrium with circular and oligomeric forms (not yet covalently linked), until this equilibrium is trapped by the ligase. This, however, does not automatically ensure that the survival probability is Boltzmann-distributed, as we show in the following.

The rate at which cyclization happens is proportional to the Boltzmann factor of the sequence

$$r_C = r_C(S) = \nu_C e^{-\beta F(S)}, \tag{13}$$

where r_C is the cyclization rate and ν_C is the attempt frequency of cyclization. DNA fragments can also be ligated to each other, causing dimerization and taking the fragments out of the pool of fragments attempting cyclization. (We are neglecting further multimerization of the dimerized fragments, which further increases the rate of dimerization of free fragments.) Assuming that the dimerization process is a second-order reaction and defining $[L]$, $[C]$, and $[D]$ as the concentrations of linear, cyclized, and dimerized fragments, respectively, and r_D as the sequence-independent rate constant for dimerization, the reaction kinetics are given by[40]

$$\frac{d[C]_S}{dt} = r_C(S)[L]_S, \tag{14}$$

$$\frac{d[D]_S}{dt} = r_D[L]_S^2, \tag{15}$$

$$\frac{d[L]_S}{dt} = -r_C(S)[L]_S - r_D[L]_S^2. \tag{16}$$

In Eqs. (14)–(16), we have explicitly written out the dependence on sequence with subscripts S. These equations hold for the concentrations of fragments with a given sequence, and we will for now only consider one sequence at a time. Therefore, in the following, we will drop the explicit subscripts.

In reality, the kinetics of fragments with different sequences are coupled because fragments may dimerize with fragments that do not have the same sequence. This means that the dimerization is much stronger than what is suggested by Eqs. (14)–(16). However, this additional dimerization is sequence-independent, and we will see that the dimerization component does not alter the qualitative behavior of the system. A qualitative characterization will be sufficient for our purposes.

This system has been treated before in the linear regime.[40] When considering different sequences with potentially very different cyclization rates, as well as different ligation times, as in the experiment of Rosanio *et al.*,[22] we may no longer be able to assume that the linear regime is valid. We therefore look for a full solution.

The probability of surviving a selection round is the probability of being cyclized at the end of the round, which is by definition

$$P(t) = \frac{[C](t)}{[C](t) + [D](t) + [L](t)} = \frac{[C](t)}{L_0}, \tag{17}$$

where L_0 is the concentration of free fragments at $t = 0$.

Equations (14)–(17) can be solved to yield

$$P(t) = -\frac{r_C}{r_D L_0}\left\{ r_C t + \log\left(\frac{r_C}{r_C + r_D L_0}\right) - \log\left(e^{r_C t} - \frac{r_D L_0}{r_C + r_D L_0}\right)\right\}. \tag{18}$$

The most important properties of this probability distribution can be understood in the limit of negligible dimerization (which can be physically achieved using a very low concentration of fragments). Without dimerization, the kinetics in Eqs. (14)–(16) simplify considerably, and Eq. (18) reduces to

$$P(t) = 1 - e^{-r_C t}, \tag{19}$$

which makes clear the saturation behavior of the probability in time. Equation (19) is plotted for different values of r_C in Fig. 3(a).

That this saturation must occur in the experiment of Rosanio *et al.*[22] follows from the values reported in Table 1 in that reference. In the last round of selection, the population consists of quickly cyclizing sequences, of which 10% is cyclized after 10 s. That means that these sequences must start to saturate within about 2 min, and hence these sequences certainly reached saturation in for instance the first selection round, which lasted 30 min.

In our model, we find that the free energies of the sequences vary over a multi-k_BT range, and as a consequence the Boltzmann factors vary over several orders of magnitude. This means that the speed with which Eq. (19) saturates to 1 also varies over several orders of magnitude. This leads to a sharp division between high-affinity and low-affinity sequences: after some time, there will be a part of the sequence population that is not undergoing selection any longer. Sequences with small enough free energy (small enough being dependent on the ligation time) all essentially have probability 1 to survive. Sequences with worse affinity are not "guaranteed" to survive and most will not be selected.

This behavior is clearly visible in the probability distributions imposed on the sequence space (determined by the free energies of the sequences), shown in Fig. 3(c). The probability distribution shows a population of sequences guaranteed to survive, a population almost guaranteed not to, and a drop-off from one to the other over a span of about 4 k_BT.

The shape of the drop-off resembles the Boltzmann distribution, as we see when we choose the cutoff time so low that no sequences saturate. In fact, in the limit $t \to 0$, we find a linear regime for Eq. (19) where the probability becomes proportional to the Boltzmann weight; unfortunately, the constant of proportionality is linear in t and therefore the efficiency of the experiment in this limit also goes to zero. (This is exacerbated by the fact that this is only true for negligible dimerization, meaning that the concentration of fragments in the experiment must be very low as well.) We may therefore hope that, apart from the lack of selection on the saturated sequences,

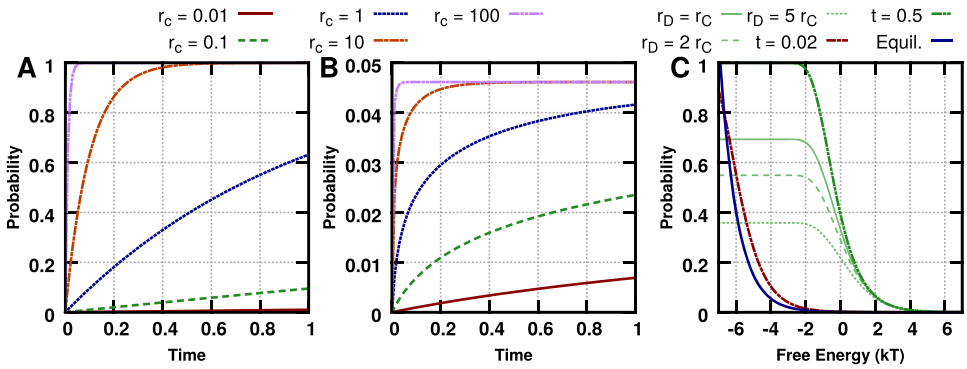

FIG. 3. Saturation behavior in the out-of-equilibrium selection method of Rosanio *et al.*[22] (a) Survival probability as a function of time, without dimerization, for values of r_C spanning several orders of magnitude. (b) As (a) with strong dimerization ($r_D/r_C = 100$). The saturation probability and how quickly it is approached change, but the overall character is similar. (c) Probability distributions imposed on the sequence space. If the probability is not allowed to saturate, the distribution (red dotted-dashed, green dotted-dotted-dashed curves) is similar but not identical to the Boltzmann distribution (blue solid curve). Also shown are the distributions for $t = 0.5$ with dimerization (light green curves), in which case the saturation probability is reduced, but the overall shape of the distribution is maintained. The free energy range is fictive, arbitrarily chosen for the purpose of illustration, but realistic.

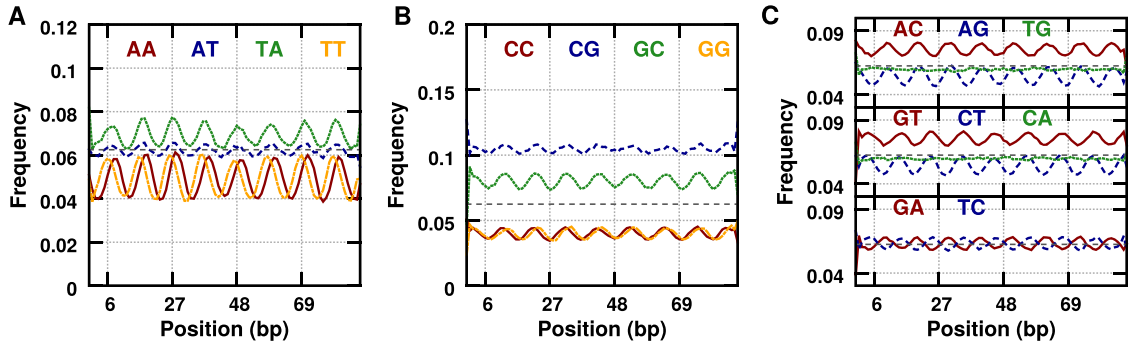

FIG. 4. Dinucleotide distributions along a ring with the Rosanio sequence, obtained from a MMC simulation at room temperature, emulating a single round of SELEX. The dinucleotides have been grouped as in Figs. 3(d)–3(f) in Ref. 22.

the selection is not qualitatively different from an equilibrium selection.

Before we show that this is the case, note that the behavior of the system in the presence of dimerization is very similar to the behavior without dimerization. In Fig. 3(b), we see that, while the saturation probability and the rapidity with which the probability approaches it are both altered by the dimerization, the overall character of the plots is similar. This is also evinced by the probability distribution in sequence space in the presence of dimerization, shown in Fig. 3(c) (light green curves). The saturation probability is different (and this is irrelevant for the competition between sequences), but the overall shape of the distribution is the same.

Let us quickly remark that the behavior we describe is actually realistic. In Fig. 3, we chose an arbitrary range of free energies to illustrate the behavior. However, we do see free energies in our model varying over roughly a range of this magnitude. More importantly, the experiment of Rosanio *et al.*[22] also evinces this behavior, as shown in Fig. 2 in that reference. This figure shows that in each round, a large percentage of fragments remain linear, meaning that in each case, the selection time was chosen such that not all sequences saturate. These reaction times vary over several orders of magnitude, and the fact that at each of these selection times a meaningful selection is taking place (the probabilities do not saturate, nor go to zero) means that the Boltzmann weights of the sequences must indeed vary over several orders of magnitude.

We must also make a remark as to the behavior of the system under multiple rounds of selection. Performing one round with time t, and one with τ, we calculate the probability to survive both rounds as the product of the probabilities to survive either round, and we find

$$P(t,\tau) = 1 - e^{-r_C t} - e^{-r_C \tau} + e^{-r_C (t+\tau)}. \tag{20}$$

If t and τ are comparable, we obtain various order terms, the lowest of which will dominate. For simplicity, assume $t = \tau$; then

$$P(t,t) = 1 - 2e^{-r_C t} + e^{-2r_C t}. \tag{21}$$

In the limit of small t, we retrieve the equilibrium statistics (by expanding the expression above to leading, i.e., second, order). If we are not in this limit (which, as explained above, is likely), the effect of the second round of selection is more subtle: the closer we are to saturation, the less effect the number of rounds has, since it only affects terms that tend to zero. In general, we find a weaker effect on the strength of the selection than in the equilibrium case [Eq. (6)].

If the ligation times of different rounds vary a lot, Eq. (20) will simply be dominated by the smallest ligation time. In that case, performing multiple rounds achieves little.

The question is how much the results of the experimental selection and our equilibrium simulation diverge. It turns out we can take the out-of-equilibrium case to an extreme, modeling it as a hard cutoff on the free energies of the system and still have a minor effect on the measured dinucleotide preferences of the system.

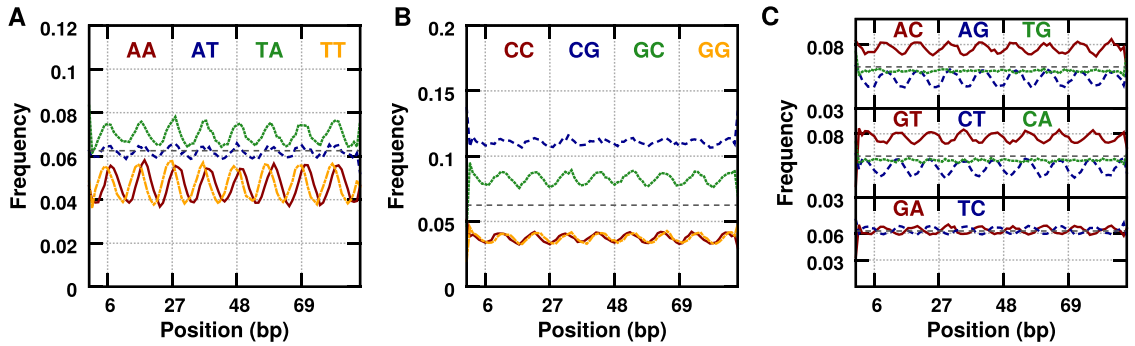

FIG. 5. Like Fig. 4, but rather than sampling according to the Boltzmann distribution, sequences were selected using a hard cutoff in the free energy [as calculated using the model from Eq. (10) and Ref. 30]. The cutoff was placed approximately at the 99th percentile of the free energies (keeping only 1%).

In order to emulate an equilibrium SELEX version of the experiment of Rosanio *et al.*,[22] we performed a MMC simulation of a closed DNA ring, modeled via the RBP model with the standard hybrid parameterization.[35,36,41] As in the SELEX experiment, we chose a ring with 126 base pairs, of which 36 were fixed to be the locking sequence from Ref. 22. The rest of the DNA was allowed to mutate. By sampling sequences during the simulation, we obtained a thermal sequence ensemble, from which we calculated the dinucleotide probability distributions shown in Fig. 4. Because we found, in Eq. (20) and onward, that the effect of multiple rounds of selection is small, we only simulated one round of selection.

We use oligonucleotide distributions calculated from the sequence ensembles as input for the approximation of Eq. (10). Using this approximation, we performed a second simulation where we generated random sequences and selected or discarded them using a hard cutoff on the free energy. The resulting dinucleotide distribution is shown in Fig. 5. We see that the calculated distributions are highly similar to each other, indicating that indeed, selecting via a Boltzmann distribution or via a hard cutoff, both lead to very similar results. Therefore, in practice, the out-of-equilibrium nature of the experiment of Rosanio *et al.* does not make for a large difference with the equilibrium scenario.

VI. RING SEQUENCE PREFERENCES *IN VITRO* AND *IN SILICO*

The dinucleotide distributions we find *in silico* show both similarities and differences with those found by Rosanio *et al.*[22] [compare Figs. 3(d)–3(f) in that reference]. First, the periodicities in the distributions, which derive from the helical nature of DNA, are very similar. The A/T-rich dinucleotides [Fig. 4(a)] are all in phase with each other, while the G/C-rich dinucleotides [Fig. 4(b)] are exactly out of phase with the former. The phasing of the other dinucleotides, shown in the three groups in Fig. 4(c), all show phasing resembling those found experimentally. For a full comparison of the phases, see Fig. 6(a).

However, one interesting deviation is the slight (1-bp) difference in phasing among the A/T-rich dinucleotides. Whereas Rosanio *et al.* find all of them peaking at exactly the same position, we find that AA generally peaks one base pair to the right of AT and TA, and TT one base pair to the left. This shift in the AA and TT dinucleotides seems to be caused by the overall preference for the TA step over the AT step. The TA step can be flanked on the left by TT but not AA and on the right by AA but not TT. This preference of the ring is analogous to the nucleosome's preference for the TTAA tetranucleotide at the positions along the nucleosome where the minor groove faces inward[32,42] and is therefore not unexpected.

The experimental distributions do not see this preference for the TA step over the AT step, which brings us to a more general difference between our theoretical results and the experimental distributions. Remarkably, the experimental probabilities never deviate very far from the uniform dinucleotide probability of 1/16. In our simulations, this is not the case: the probabilities take on values from around 0.04, up to around 0.12. This does not occur only locally, but several dinucleotides have an average probability, along the entire ring, significantly different from the uniform value, see Fig. 6(b).

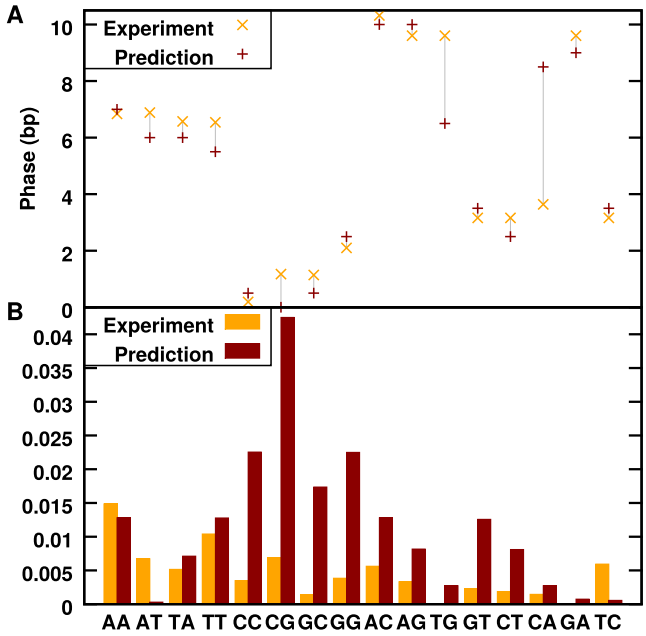

FIG. 6. Comparison between the experimentally found sequence preferences of the DNA ring and those predicted by our model. (a) The relative phases of the oscillatory signals. Here the experimental results and theoretical predictions generally agree to within one base pair. (For TG and CA, where the phases do not match, our model predicts no discernible oscillations, and hence the phase is poorly defined.) (b) The absolute deviations of the average probabilities of the sixteen dinucleotides from the uniform value 1/16. The experimental probabilities mostly oscillate closer to the uniform value than the theoretical ones.

The uniformity of the experimentally obtained distributions is surprising. It is known, for instance, that the affinity of nucleosomes to sequences correlates with GC content. Therefore, e.g., the enrichment we observe of the CG and GC dinucleotides is not unexpected. More generally, there is no reason to expect all the dinucleotides to have probabilities close to 1/16.

This discrepancy between theoretical prediction and experiment could have several reasons. It may be a failure of the RBP model or its parameterization, which have been tested most extensively in the context of nucleosomes. It is possible that rings are less similar to nucleosomes than one might expect and that the model does not capture the difference. For example, Rosanio *et al.* find longer-range correlations in their sequences. The RBP model contains such interactions only indirectly, due to the thermal nature of the system and the constraints placed on the DNA, but microscopically only accounts for nearest-neighbor interactions. There is much evidence that the RBP model is an oversimplification in this regard.[43–48]

Other potential causes exist on the experimental side. For instance, the experiment of Rosanio *et al.* employed different ionic conditions than those that are generally used for nucleosome reconstitution experiments. Ionic conditions are known to affect DNA elasticity.[49] In particular, Rosanio *et al.* used a significant concentration of magnesium, whose ions

are known to strongly affect DNA mechanics.[50,51] Such differences in experimental conditions may contribute to the observed differences.

Another potential suspect is the fact that the experimental selection process only had access to a limited set of sequences, whereas the MMC algorithm can access all of sequence space. We might expect this to limit the ability of the experiment to select for coherent sequence properties like GC content and correctly phased dinucleotides. However, this possible cause is ruled out by the results presented in Fig. 5. It shows the dinucleotide preferences of high-affinity sequences selected from a pool of only $\sim 10^7$ random sequences, far smaller than the library constructed experimentally (10^{13} sequences). However, the results still exhibit the same non-uniform dinucleotide probabilities. Therefore, limited pool size cannot explain the differences we find.

In conclusion, the uniformity in the average probability found by Rosanio *et al.* is not currently understood from a DNA-mechanical point of view and more research is needed to understand where the difference between DNA rings and nucleosomes originates.

VII. SELEX SIMULATION FOR SMALL AND OVERWOUND CIRCLES

A further benefit of our ability to perform SELEX experiments *in silico* is that it allows for studying systems that are experimentally difficult to realize, such as very small rings, rings whose length is not an integer multiple of the helical period of DNA, or overwound rings. These all have a high energetic cost and are therefore slow to form, as they are dependent on thermal fluctuations for ligation. In our simulations, we can simply impose the desired constraints from the beginning, and we do not need to wait for the system of interest to form spontaneously.

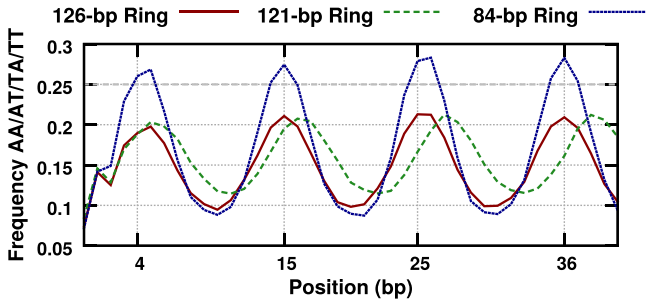

FIG. 7. The total frequency of the A/T-rich dinucleotides (AA/AT/TA/TT) for three different rings, all directionally biased using the artificial locking sequence from Ref. 34 (see Sec. V): the 126-base-pair ring considered before (red solid curve), a 121-base-pair ring, which requires over- or undertwisting of the DNA (dashed green curve) and a significantly shorter (but not overwound) 84-base-pair ring (dotted blue curve). All three curves were calculated at room temperature, with the mutation temperature reduced to 1/3 of room temperature. This was achieved as described in Sec. III. A ring whose length is not an integer multiple of the helical repeat stretches (in this case, where the ring is underwound) the periodicity of the distributions and slightly reduces their amplitude. A tighter ring leads to larger amplitudes.

Figure 7 presents a part of the AA/AT/TA/TT dinucleotide distributions for three different rings: the 126-base-pair ring analogous to the one used by Rosanio *et al.*, a slightly shorter, 121-base-pair ring (which leads to a slightly twisted ring because the length is not an integer multiple of the helical period), and a much shorter 84-base-pair ring, with correspondingly larger curvature. All rings are direction-biased not using the locking sequence of Rosanio *et al.*, but with the artificial, strongly bent sequence from Ref. 34, described in Sec. V. We chose this sequence over the locking sequence from

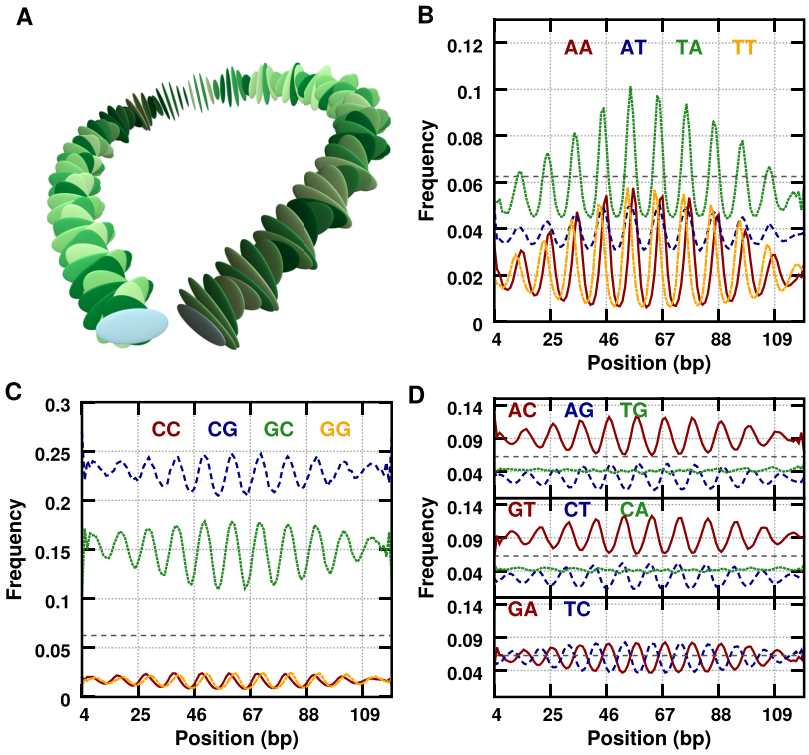

FIG. 8. (a) A teardrop-shaped DNA molecule, held in place at its ends. (b)–(d) Dinucleotide distributions calculated for the teardrop-shaped DNA, at room temperature, with the mutation temperature reduced to 1/3 of room temperature. The teardrop is most strongly curved in the center and more straight toward the ends of the molecule. This leads to distributions similar to those of rings, but whose amplitudes taper off toward the ends.

Rosanio *et al.* because of the stronger and cleaner directional bias (see Fig. 2).

The 84-base-pair ring is more tightly curved and therefore places a stronger selection on the sequences, leading to the higher amplitude in the frequencies. For ease of comparison, Fig. 7 only shows the combined frequencies of the A/T-rich dinucleotides, but the same effect applies to all individual dinucleotide frequencies.

The 121-base-pair ring is underwound by half a turn, and the periodicity in the dinucleotide frequencies is correspondingly stretched to a slightly larger period. The amplitude is not increased, as we found when shortening the ring to 84 base pairs, but is rather slightly decreased. This is in fact as expected: the locking sequence becomes less effective when the DNA is underwound because it is designed to give coherent curvature in unconstrained DNA. The twist mismatch weakens the directional bias imparted by the locking sequence. As for the 84-base-pair ring, these observations are conserved among all dinucleotide probabilities, not only those shown in Fig. 7.

We could underwind or overwind our rings by more than half a turn, and we would expect similar stretching and compression of the periodic nature of the frequencies. However, we will start to run into two complications. First, as already observed, the locking sequence will become less effective. (We could design locking sequences specifically for overwound or underwound DNA, but that is beyond the scope of the current work.) Second, for strongly overwound or underwound DNA, it will become energetically favorable to supercoil.[52,53] This complicates the system because different parts of the DNA will interact and steric interactions must be taken into account.

We modeled half of a figure-eight supercoil of DNA as a simple teardrop shape as a proof-of-principle. This model consists of two constraints: we place the base pairs of our molecule along a teardrop-shaped curve and keep the first and last base pairs fixed throughout the simulation. An example state is shown in Fig. 8(a). Such a shape, although it is essentially two-dimensional and therefore a simplification of real three-dimensional supercoiling configurations, emulates the basic geometry of the end-loops of supercoils[54] and protein-induced DNA loops.[55,56]

Applying our methodology to such a teardrop shape, consisting of 126 base pairs, we find the dinucleotide frequencies presented in Figs. 8(b)–8(d). As expected, the distributions we find resemble those of a ring. However, because the curvature is not constant—it falls off toward the ends of the molecule—the amplitude of the distributions tapers off.

VIII. CONCLUSION

We have presented methods to emulate, *in silico*, equilibrium SELEX experiments. The MMC method[32] is akin to such experiments and can be used to select for high-affinity sequences for a given DNA system. One limitation of the MMC method was that the selection pressure on the sequences and the temperature in the simulation are linked. In an equilibrium SELEX experiment, the mutation pressure is modified in a mathematically straightforward way by the number of rounds of selection applied.

We employed the methodology of Ref. 30, which makes use of the output of a MMC simulation to build a model for sequence-dependent nucleosome affinity, to resample sequence space at a different mutation temperature, without altering the temperature employed for the spatial moves. This separation of mutation pressure and physical temperature allows us to more closely mimic the outcome of a SELEX experiment, as well as learn more about our systems in general.

We have used this new methodology to examine various systems. First, in Sec. IV, we assessed how changing the physical temperature, without changing the mutation pressure, affects the sequence preferences of nucleosomes and rings. We found that, due to the strongly constrained nature of the nucleosome, entropic contributions to the free energy do not play an influential role, and consequently the sequence preferences of the nucleosome are not strongly temperature-dependent (in the range between 1/4 of room temperature and room temperature). Rings, on the other hand, are not heavily constrained systems, which means that the entropic contribution to their free energy is more important and the sequence preferences of rings depend strongly on temperature.

In Sec. V, we considered the SELEX experiment for rings performed by Rosanio *et al.*[22] This experiment is not an ideal equilibrium SELEX experiment because it uses irreversible reactions, and we examined what effect this has on the (non-Boltzmann) distribution the experiment imposes on sequence space. While some differences in the methodology must be noted, the effects on the measured sequence preferences turned out to be small and we were able to compare the predictions of our *in silico* SELEX experiment for rings with the experimental results. We found that the periodic nature of the dinucleotide distributions in rings is well captured by the RBP model we employed to model the DNA. However, some differences are apparent, the most striking one being that we predict significant deviation away from 1/16 in the overall frequencies of dinucleotides. For example, we find the CG dinucleotide significantly enriched, similar to what we find for nucleosomes.

The experimental distributions show very little overall variation away from 1/16, meaning that no dinucleotides are significantly enriched or depleted along the entire length of the ring. As discussed in Sec. VI, this difference might point to a failure of the RBP model or its parameterization to capture an unknown difference between rings and nucleosomes. It is also possible that the experimental conditions caused a change in the behavior of the DNA, leading to the theoretically unexpected difference. Whichever the case, more research is needed to answer this question.

We finally applied our methods to several systems that would be difficult to access experimentally. We considered rings that would have difficulty forming because they either consist of only a short piece of DNA, requiring tight curvature, or because their length is not an integer multiple of the helical repeat length of DNA. We also considered a teardrop-shaped DNA molecule, which mimics a part of strongly overwound (or underwound) DNA, or a protein-induced antiparallel DNA loop. Despite being difficult to reproduce in the lab, DNA that is sharply bent into circular or antiparallel loops plays

important roles in biology,[57] and we showed that our methods can be used to determining the sequence preferences of such systems, opening up new possibilities of examining systems that have been inaccessible until now.

The methodology we have presented relies on a sequence-dependent description of DNA mechanics, for which role we have cast the rigid base-pair model. However, the methods are general and can be used with any other underlying model, and they are straightforward to update if and when more advanced DNA models become available in the future.

ACKNOWLEDGMENTS

We would like to thank Ralf Everaers for useful discussions. This work was supported by the Netherlands Organisation for Scientific Research (NWO/OCW), as part of the Frontiers of Nanoscience program.

[1] R. Stoltenburg, C. Reinemann, and B. Strehlitz, Biomol. Eng. **24**, 381 (2007).
[2] J. D. Kahn and D. M. Crothers, Biochem. **89**, 6343 (1992).
[3] J. D. Parvin, R. J. McCormick, P. A. Sharp, and D. E. Fisher, Nature **373**, 724 (1995).
[4] T. Schätz and J. Langowski, J. Biomol. Struct. Dyn. **15**, 265 (1997).
[5] N. A. Davis, S. S. Majee, and J. D. Kahn, J. Mol. Biol. **291**, 249 (1999).
[6] L. Bracco, D. Kotlarz, A. Kolb, S. Diekmann, and H. Buc, EMBO J. **8**, 4289 (1989).
[7] M. R. Gartenberg and D. M. Crothers, J. Mol. Biol. **219**, 217 (1991).
[8] J. Pérez-Martín and V. de Lorenzo, Annu. Rev. Microbiol. **51**, 593 (1997).
[9] H. M. Wu and D. M. Crothers, Nature **308**, 509 (1984).
[10] H. Yamada, S. Muramatsu, and T. Mizuno, J. Biochem. **108**, 420 (1990).
[11] G. Prosseda, M. Falconi, M. Giangrossi, C. O. Gualerzi, G. Micheli, and B. Colonna, Mol. Microbiol. **51**, 523 (2004).
[12] T. E. Cloutier and J. Widom, Proc. Natl. Acad. Sci. U. S. A. **102**, 3645 (2005).
[13] J. Wei, L. Czapla, M. A. Grosner, D. Swigon, and W. K. Olson, Proc. Natl. Acad. Sci. U. S. A. **111**, 16742 (2014).
[14] A. R. Cutter and J. J. Hayes, FEBS Lett. **589**, 2914 (2015).
[15] M. Radman-Livaja and O. J. Rando, Dev. Biol. **339**, 258 (2010).
[16] B. Eslami-Mossallam, H. Schiessel, and J. van Noort, Adv. Colloid Interface Sci. **232**, 101 (2016).
[17] H. R. Widlund, H. Cao, S. Simonsson, E. Magnusson, T. Simonsson, P. E. Nielsen, J. D. Kahn, D. M. Crothers, and M. Kubista, J. Mol. Biol. **267**, 807 (1997).
[18] H. Cao, H. R. Widlund, T. Simonsson, and M. Kubista, J. Mol. Biol. **281**, 253 (1998).
[19] P. T. Lowary and J. Widom, J. Mol. Biol. **276**, 19 (1998).
[20] K. A. Bailey, S. L. Pereira, J. Widom, and J. N. Reeve, J. Mol. Biol. **303**, 25 (2000).
[21] B. A. Beutel and L. Gold, J. Mol. Biol. **228**, 803 (1992).
[22] G. Rosanio, J. Widom, and O. C. Uhlenbeck, Biopolymers **103**, 303 (2015).
[23] B. S. Singer, T. Shtatland, D. Brown, and L. Gold, Nucleic Acids Res. **25**, 781 (1997).
[24] S. C. Satchwell, H. R. Drew, and A. A. Travers, J. Mol. Biol. **191**, 659 (1986).
[25] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I. K. Moore, J.-P. Z. Wang, and J. Widom, Nature **442**, 772 (2006).
[26] Y. Field, N. Kaplan, Y. Fondufe-Mittendorf, I. K. Moore, E. Sharon, Y. Lubling, J. Widom, and E. Segal, PLoS Comput. Biol. **4**, e1000216 (2008).
[27] N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, Y. Field, E. M. LeProust, T. R. Hughes, J. D. Lieb, J. Widom, and E. Segal, Nature **458**, 362 (2009).
[28] K. Brogaard, L. Xi, J.-P. Wang, and J. Widom, Nature **486**, 496 (2012).
[29] G. Moyle-Heyrman, T. Zaichuk, L. Xi, Q. Zhang, O. C. Uhlenbeck, R. Holmgren, J. Widom, and J.-P. Wang, Proc. Natl. Acad. Sci. U. S. A. **110**, 20158 (2013).
[30] M. Tompitak, G. T. Barkema, and H. Schiessel, BMC Bioinf. **18**, 157 (2017).
[31] M. Tompitak, C. Vaillant, and H. Schiessel, Biophys. J. **112**, 505 (2017).
[32] B. Eslami-Mossallam, R. D. Schram, M. Tompitak, J. van Noort, and H. Schiessel, PLoS ONE **11**, e0156905 (2016).
[33] M. Tompitak, L. de Bruin, B. Eslami-Mossallam, and H. Schiessel, Phys. Rev. E **95**, 052402 (2017).
[34] M. Tompitak, H. Schiessel, and G. T. Barkema, EPL **116**, 68005 (2016).
[35] W. K. Olson, A. A. Gorin, X. J. Lu, L. M. Hock, and V. B. Zhurkin, Proc. Natl. Acad. Sci. U. S. A. **95**, 11163 (1998).
[36] N. B. Becker, L. Wolff, and R. Everaers, Nucleic Acids Res. **34**, 5638 (2006).
[37] L. de Bruin, M. Tompitak, B. Eslami-Mossallam, and H. Schiessel, J. Phys. Chem. B **120**, 5855 (2016).
[38] I. M. Kulić, R. Thaokar, and H. Schiessel, EPL **72**, 527 (2005).
[39] I. M. Kulić, R. Thaokar, and H. Schiessel, J. Phys.: Condens. Matter **17**, S3965 (2005).
[40] W. H. Taylor and P. J. Hagerman, J. Mol. Biol. **212**, 363 (1990).
[41] F. Lankaš, J. Sponer, J. Langowski, and T. E. Cheatham III, Biophys. J. **85**, 2872 (2003).
[42] C. K. Collings, A. G. Fernandez, C. G. Pitschka, T. B. Hawkins, and J. N. Anderson, PLoS ONE **5**, e10933 (2010).
[43] K. Yanagi, G. G. Privé, and R. E. Dickerson, J. Mol. Biol. **217**, 201 (1991).
[44] M. J. Packer, M. P. Dauncey, and C. A. Hunter, J. Mol. Biol. **295**, 85 (2000).
[45] E. J. Gardiner, C. A. Hunter, M. J. Packer, D. S. Palmer, and P. Willett, J. Mol. Biol. **332**, 1025 (2003).
[46] S. B. Dixit, D. L. Beveridge, D. A. Case, T. E. Cheatham III, E. Giudice, F. Lankaš, R. Lavery, J. H. Maddocks, R. Osman, H. Sklenar, K. M. Thayer, and P. Varnai, Biophys. J. **89**, 3721 (2005).
[47] S. Fujii, H. Kono, S. Takenaka, N. Go, and A. Sarai, Nucleic Acids Res. **35**, 6063 (2007).
[48] R. Lavery, K. Zakrzewska, D. L. Beveridge, T. C. Bishop, D. A. Case, T. E. Cheatham III, S. B. Dixit, B. Jayaram, F. Lankaš, C. Laughton, J. H. Maddocks, A. Michon, R. Osman, M. Orozco, A. Perez, T. Singh, N. Spackova, and J. Sponer, Nucleic Acids Res. **38**, 299 (2009).
[49] L. D. Williams and L. J. Maher III, Annu. Rev. Biophys. Biomol. Struct. **29**, 497 (2000).
[50] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, Proc. Natl. Acad. Sci. U. S. A. **94**, 6185 (1997).
[51] M. Guéroult, O. Boittin, O. Mauffret, C. Etchebest, and B. Hartmann, PLoS ONE **7**, e41704 (2012).
[52] A. V. Vologodskii, S. D. Levene, K. V. Klenin, M. Frank-Kamenetskii, and N. R. Cozzarelli, J. Mol. Biol. **227**, 1224 (1992).
[53] A. Fathizadeh, H. Schiessel, and M. R. Ejtehadi, Macromolecules **48**, 164 (2015).
[54] M. Emanuel, G. Lanzani, and H. Schiessel, Phys. Rev. E **88**, 022706 (2013).
[55] A. Balaeff, L. Mahadevan, and K. Schulten, Phys. Rev. E **73**, 031919 (2006).
[56] I. M. Kulić, H. Mohrbach, R. Thaokar, and H. Schiessel, Phys. Rev. E **75**, 011913 (2007).
[57] R. Schleif, Annu. Rev. Biochem. **61**, 199 (1992).