Mesoscopic modeling of DNA denaturation rates: Sequence dependence and experimental comparison

Oda Dahlen, and Titus S. van Erp

Articles you may be interested in

DNA denaturation bubbles: Free-energy landscape and nucleation/closure rates
The Journal of Chemical Physics **142**, 034903 (2015); 10.1063/1.4905668

Model simulations of DNA denaturation dynamics
The Journal of Chemical Physics **114**, 579 (2001); 10.1063/1.1329137

Sufficient minimal model for DNA denaturation: Integration of harmonic scalar elasticity and bond energies
The Journal of Chemical Physics **145**, 144101 (2016); 10.1063/1.4964285

Thermal and mechanical denaturation properties of a DNA model with three sites per nucleotide
The Journal of Chemical Physics **135**, 085105 (2011); 10.1063/1.3626870

Sequence-dependent theory of oligonucleotide hybridization kinetics
The Journal of Chemical Physics **140**, 175104 (2014); 10.1063/1.4873585

Coarse-grained modeling of DNA oligomer hybridization: Length, sequence, and salt effects
The Journal of Chemical Physics **141**, 035102 (2014); 10.1063/1.4886336
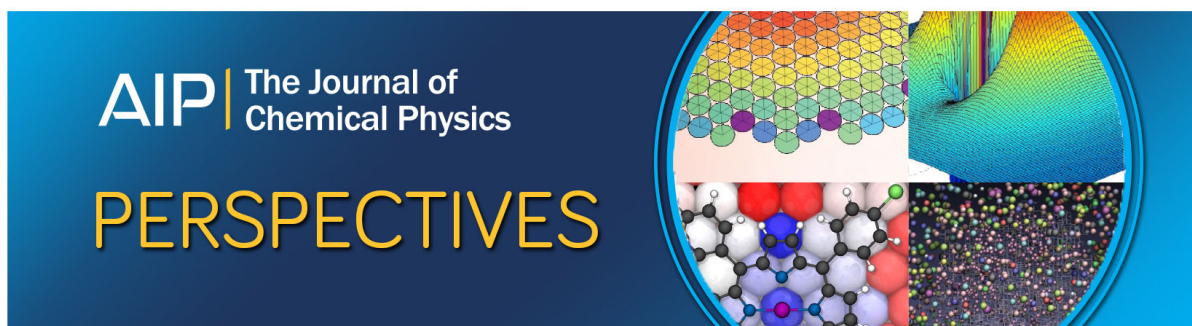
Mesoscopic modeling of DNA denaturation rates: Sequence dependence and experimental comparison

Oda Dahlen[a) and Titus S. van Erp[b)
Department of Chemistry, Norwegian University of Science and Technology (NTNU), Høgskoleringen 5, Realfagbygget D3-117 7491 Trondheim, Norway

Using rare event simulation techniques, we calculated DNA denaturation rate constants for a range of sequences and temperatures for the Peyrard-Bishop-Dauxois (PBD) model with two different parameter sets. We studied a larger variety of sequences compared to previous studies that only consider DNA homopolymers and DNA sequences containing an equal amount of weak AT- and strong GC-base pairs. Our results show that, contrary to previous findings, an even distribution of the strong GC-base pairs does not always result in the fastest possible denaturation. In addition, we applied an adaptation of the PBD model to study hairpin denaturation for which experimental data are available. This is the first quantitative study in which dynamical results from the mesoscopic PBD model have been compared with experiments. Our results show that present parameterized models, although giving good results regarding thermodynamic properties, overestimate denaturation rates by orders of magnitude. We believe that our dynamical approach is, therefore, an important tool for verifying DNA models and for developing next generation models that have higher predictive power than present ones. © 2015 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4922519]

I. INTRODUCTION

Gaining knowledge and understanding of the structure and function of DNA, our genetic material, is crucial for dealing with diseases related to DNA. In 2004, the mapping of the complete human genome was accomplished,[1] leading to a better understanding of DNA-related diseases. However, knowledge of the DNA structure is not enough to sufficiently understand biological processes.[2] For this, we also need to understand how the structure affects the equilibrium properties and the dynamics of the DNA molecule. In fundamental genetic processes, such as transcription and replication, DNA must undergo dynamical changes.[3] Due to their high degree of complexity, the details of transcription and replication are not adequately understood and a satisfactory descriptive model is difficult to design. However, these processes initiate with the formation of local "denaturation" bubbles in a similar way as the process of DNA denaturation, or DNA melting, which is a considerably simpler process to study theoretically and experimentally. Thus, studying DNA denaturation is, in addition to being very interesting in itself, considered a well-grounded step towards the full comprehension of the mechanisms involved in transcription and replication.

There are two main objectives of this article. First, we extend the study of Ref. 4 by examining the denaturation rate constant as a function of a wider range of sequences and temperatures. Second, we look at a simple adaption of the PDB model to study hairpins for which experimental denaturation rates are available.

II. THE PEYRARD-BISHOP-DAUXOIS (PBD) MODEL

The PBD model is a well-known mesoscopic model for describing the thermodynamics and dynamics of DNA. It simplifies the geometry of the DNA double helix by describing it as a one-dimensional chain, where each base pair (bp) is represented as a point-mass along the chain. A DNA chain of N bps has potential energy $U(\{y_i\}) = V_1(y_1) + \sum_{i=2}^{N} V_i(y_i) + W(y_i, y_{i-1})$, where

$$V_i(y_i) = D_i(e^{-a_i y_i} - 1)^2, \tag{1}$$

$$W(y_i, y_{i-1}) = \frac{1}{2}K(1 + \rho e^{-\alpha(y_i + y_{i-1})})(y_i - y_{i-1})^2. \tag{2}$$

The transverse displacements of the nucleotides from equilibrium are denoted y_i, i being the bp-index. V_i is the Morse potential, describing the repulsive interaction due to the hydrogen bonds between two bases on opposite strands, the repulsion of phosphate groups, as well as the screening by the surrounding solvent. The depth of the potential is represented by D_i and the width by a_i. The depth D_s and the width a_s for strong GC bps are larger than the depth D_w and the width a_w for weak AT bps. W represents the stacking potential that describes the interaction between neighboring bps and includes a harmonic and a nonlinear term. The term $\rho e^{-\alpha(y_i + y_{i-1})}$ was introduced in the improvement of the original PB-model,[5] leading to the PBD model,[6] and this term is the reason why the PBD model can reproduce sharp phase transitions observed in denaturation experiments.[7-11] If one of the bps stretches far beyond α^{-1}, $\rho e^{-\alpha(y_i + y_{i-1})} \approx 0$, and the effective strength of the potential drops from $K(1 + \rho)$ to K, which resembles the decline in the overlapping π-electrons occurring as one of two neighboring bases breaks and moves out of stack.

There are of course also some limitations that should be considered. (i) The PBD model does not distinguish between

---

a)Electronic address: oda.dahlen@ntnu.no
b)Electronic address: titus.van.erp@ntnu.no

A and T bases or between G and C bases. In other words, the AT bp has the same physics as the TA bp and the same is true for the GC and CG bps. (ii) The stacking potential $W(y_i, y_{i-1})$ depends on the positions of the bps i and $i - 1$ but not on their identity. (iii) The model does not contain helical stress. (iv) The PBD model includes solvent effects only implicitly via the effective interaction parameters. (v) The one-dimensional character of the model becomes less trustable at very large bp separations. Despite these approximations, several studies have shown that, after fitting of the PBD parameters, denaturation experiments can be reasonably well described.[7,9] In addition, the PBD model has also been shown to reproduce specific features seen in experiments, such as bubble formation in DNA strands[12] and the previously mentioned sharp phase transitions observed in denaturation experiments.[10] A strong point of the PBD model is that it is very fast. Molecular dynamics (MD) simulations of long DNA chains up to several hundreds of bps can be run up to the microsecond time scale which would not be feasible for full-atom simulations. Yet, the model has more detail than Ising-type models such as the Poland-Scheraga model.[13] This allows studying more realistic dynamics. Besides computational efficiency, the simplicity of the model has another advantage. It allows gaining intuition and insight that is more difficult to get from a full-atomistic simulation if it would be feasible. A minimal model that is describing the experimental features is, therefore, extremely valuable.

However, due to a lack of experimental evidence, it is yet uncertain whether present parameterized PBD models can really give qualitative correct denaturation dynamics or that extensions are required. Two parameter sets frequently used for the PBD-model are the one by Campa and Giansanti[7,8] having $K = 0.025$ eV/Å^2, $\rho = 2$, $\alpha = 0.35$ Å^{-1}, $D_w = 0.05$ eV, $D_s = 0.075$ eV, $a_w = 4.2$ Å^{-1}, and $a_s = 6.9$ Å^{-1}, and the more recent set by Theodorakopoulos,[9] where $K = 0.00045$ eV/Å^2, $\rho = 50$, $\alpha = 0.20$ Å^{-1}, $D_w = 0.1255$ eV, $D_s = 0.1655$ eV, $a_w = 4.2$ Å^{-1}, and $a_s = 6.9$ Å^{-1}. In the following, we will call these parameter sets PBD-CaGi and PBD-Th, respectively. The most noticeable differences between these two parameter sets is the value of the nonlinear stacking parameter, ρ, and the strength of the potential, K. The large ρ-value was advocated by Theodorakopoulos based on the large differences in stiffness of single stranded and double stranded DNA (dsDNA).[9] The effective coupling constants with two consecutive bps in stack $K(1 + \rho)$ are closer but still deviating by more than a factor of 3: 0.075 eV/Å^2 versus 0.023 eV/Å^2. The two parameter sets are based on completely different physics with the DNA model PBD-Th being much more flexible than PBD-CaGi, but they both give reasonable results when compared to data from actual denaturation experiments. This shows that there is a need for additional experimental and theoretical data to verify mesoscopic DNA models.

III. PBD ADAPTATION FOR HAIRPINS

A DNA hairpin is a secondary structure of DNA and RNA, found most frequently in RNA, where it is one of the entities of many RNA configurations. A hairpin consists of a bp stem, which is a dsDNA chain, and a base loop at the end of the stem,

which is a single stranded DNA chain. Hairpins are highly dynamic structures, and in a simplified definition, we can say that they fluctuate between two main states, the closed state, where the bps of the stem are paired, and the open state, where the bps of the stem are free. Hairpins are vital for several of the functions of DNA and RNA. They are involved in regulation of the transcription by binding to proteins,[14] the regulation of gene expression,[15] and intermediary hairpin structures participate in replication and recombination.[16–18] With respect to DNA based molecular nano-devices, DNA/RNA hairpins have in particular been proposed suitable for storing molecular memory[19] and as engines to drive nano-devices.[20,21]

Two possible ways to adapt the PBD model for studying DNA hairpins have been proposed. Errami et al.[22] used the PBD model for describing the stem in combination with the Kratky-Porod model for describing the loop. For reasons of computational efficiency, however, we adopted the simpler approach by Hanne et al.[23] in which the flexibility of bp at the end of the stem, just before the loop starts, is restricted by an additional confining potential that is added to the Morse potential. In our case, we applied an exponential potential to the terminal bp of the dsDNA stem to mimic the "hairpin-effect,"

$$V(y_N) = \begin{cases} \dfrac{\text{eV}}{\text{Å}^6}(y_N - \tau)^6, & \text{when} \quad y_N > \tau \\ 0, & \text{when} \quad y_N < \tau \end{cases}. \tag{3}$$

Naturally, τ should in principle depend on the actual loop length of the hairpin since that determines the maximum separation of this bp. However, also steric hindrance and interactions with solvents will be important.

IV. RATE CONSTANTS

Despite the fact that mesoscopic character of the PBD model allows running much longer simulations than one can achieve with full-atom models, it is generally still not possible to study denaturation of long DNA polymers at temperatures that are significantly lower than the denaturation temperature using MD. The reason is that denaturation is a rare event on the time scale that can be achieved by brute force MD, even for such a simplified model. Henceforth, in order to study the denaturation transition, we need to invoke special *rare event* simulation techniques. Several simulation techniques have been developed that are able to reproduce the exact same rate constant as brute force MD, but orders of magnitude faster. The Reactive Flux (RF) method is one of them, developed last century by Eyring,[24] Wigner,[25] and Keck[26] based on the concepts of a transition state dividing surface, the free energy as a function of a reaction coordinate, and the transmission coefficient. The transmission coefficient corrects for the transition state theory (TST) expression. There are several methods to calculate the free energy and the transmission coefficient (See, e.g., Refs. 27 and 28). Alternative methods are based on transition path sampling (TPS)[29] which became an effective method for calculating rate constant when the transition interface sampling (TIS) algorithm[30] was developed. Popular methods based on variations of the TIS algorithm are forward

flux sampling (FFS)[31] and replica exchange TIS (RETIS).[4] Both these methods are in principle exact methods though FFS can have severe sampling problems with transition paths not moving along the most likely pathway.[32] RETIS can be more efficient than the reactive flux method in complex systems when hysteresis[27] can hamper the standard free energy calculation methods and, in addition, many trajectories are needed to determine small transmission coefficients. For the PBD model, however, the RF method is by far the most efficient method due to the fact that it only has first-neighbor interactions which allows an extremely efficient and accurate free energy calculation method. In this method, the multidimensional partition-function integrals are solved in an iterative algorithm.[33] The free energy barrier can, hence, be determined without any statistical error. Only a small systematic error due to the step-size in the numerical integration scheme remains, but overall the free energy barrier can be determined up to several digits. Giving this beneficial approach for determining the free energy, the overall efficiency of the method is very good even if the transmission coefficient is rather low (for all sequences around 0.02).

The RF method writes the rate constant as the following product: $k = \kappa k^{\text{TST}}$. Here, k^{TST} is the transition theory expression for the rate constant that only requires information of the free energy profile along the reaction coordinate.[27] κ is the correction to the TST expression and always has a value between 0 and 1. This factor corrects for fast correlated recrossing and can be calculated by releasing many trajectories from the top of the free energy barrier. Although several expressions have been used for calculating κ, the one based on the *effective positive flux* expression[34,35] is probably the most efficient and used in this work. Thus, the rate constant calculation was obtained in a two-step procedure. First, the free energy was calculated along the reaction coordinate λ which was chosen as $\lambda = \text{MIN}(\{y_i\})$ just like in previous publications.[4,33] The integration of the partition function integrals was carried out using the iterative direct integration method.[4,12,33,36] The step-size for the numerical integration was set to $dy = 0.05$ Å and the integration boundaries were set according to Ref. 12 using a tolerance of $\epsilon = 10^{-40}$. The transmission coefficients were obtained by running 10^6 trajectories, using a time step of 1 fs and bp masses of 300 amu. In all simulations, Langevin dynamics were applied. The friction coefficient, γ, was set to 50 ps^{-1}, which is considered a suitable value for the friction of water.[37]

Experimental denaturation rate constants are mainly available from DNA-hairpins. In Ref. 38, the authors obtained kinetic and thermodynamic data on folding and unfolding of individual DNA hairpins. The hairpins were unzipped and re-folded at constant force by using an ultrastable, high-precision optical force clamp. The two ends on the hairpin were attached to dsDNA handles bound to polystyrene beads held in two independently controlled laser traps. The unfolding rates were measured by monitoring the extension of the molecule as the hairpin folded and unfolded under constant force. In Ref. 23, the authors performed single-molecule measurements of the opening rate of DNA hairpins under mechanical stress. The opening rate for zero force is obtained by measuring how the opening rate depends on a applied external force field acting on

the hairpin. Bonnet *et al.*[39] determined experimental reaction rates using a combination of fluorescence energy transfer and fluorescence correlation spectroscopy.

V. RESULTS AND DISCUSSION

A. dsDNA chains

We investigated DNA chains of lengths ranging from 4 to 100 bps with a GC-content of 1/3, 1/2, and 2/3. For each of these chains, we examined how the order of the sequence, i.e., the distribution of weak and strong bps, influenced the denaturation rate. We considered four different ways to order the sequences which we will denote "AAGG"-, "AGGA"-, "GAAG"-, and "AGAG"-sequences. This shorthand notation refers to 4 bp sequence having that order. For larger sequences, this notation is explained in Table I. In words, the "AGGA"-chain is the chain with a strong block in the middle, the "GAAG"-chain is the one with the weak block in the middle, the "AAGG"-chain has the weak and strong bps at either side, and finally "AGAG" is the alternating chain, having the most even distribution of weak and strong bps. Figure 1 shows the rate constant of the "AGAG" chain with 33%, 50%, and 66% GC-content for the PBD-CaGi parameters. The results show that at low temperatures, the shortest chains have the highest denaturation rates, whereas at the high temperatures, it is reverse with the longer chains denaturing faster. The transition between the two regimes seems to occur at the critical melting temperature, i.e., where the denaturation phase transition occurs in the infinite chain limit (325 K and 365 K for homogeneous AT and GC, respectively[12]). The same trend was found in Fig. 2 of Ref. 33, where the denaturation constant as a function of chain-length was plotted for different temperatures in the range of 300–350 K. This could seem surprising since one might intuitively expect that the longer chains open up more slowly than the short ones at all times. Apparently, that is not the case at temperatures above the melting temperature. This can be explained as follows. Since a denaturation event begins with openings of single bps, which develops to

TABLE I. Explanation of the notation used in this article to denote a certain bp ordering. "AAGG," "AGGA," "GAAG," and "AGAG" refer to the sequences of four bps having this specific order. In the right column, we show what we mean by this order for a 12-bps sequence having either 1/3, 1/2, or 2/3 of the bps being GC.

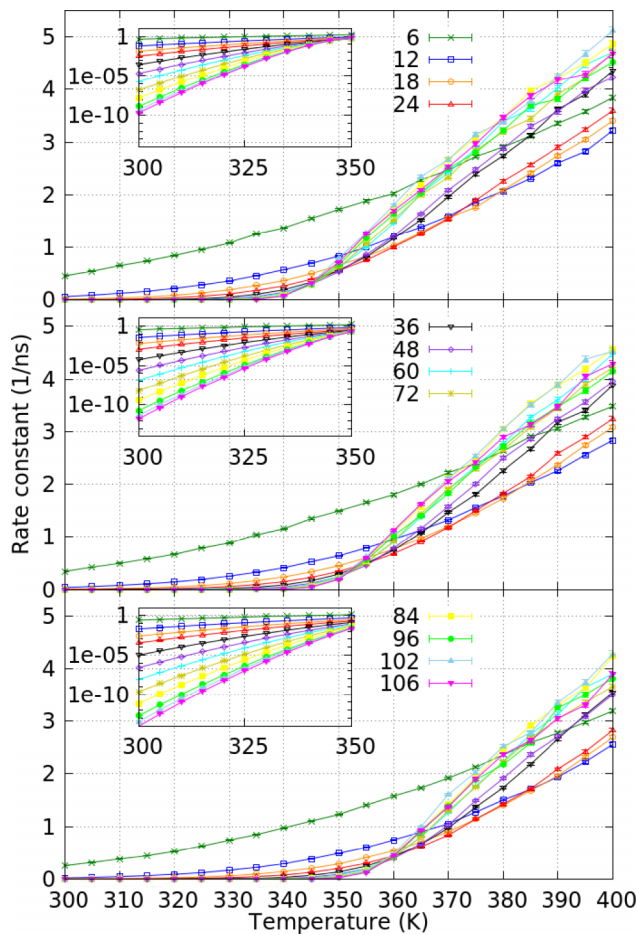| Sequence notation	Percentage GC-bps (%)	Sequences of 12 bps
"AAGG"	33	AAAAAAAAGGGG
"AAGG"	50	AAAAAAGGGGGG
"AAGG"	66	AAAAGGGGGGGG
"AGGA"	33	AAAAGGGGAAAA
"AGGA"	50	AAAGGGGGGAAA
"AGGA"	66	AAGGGGGGGGAA
"GAAG"	33	GGAAAAAAAAGG
"GAAG"	50	GGGAAAAAAGGG
"GAAG"	66	GGGGAAAAGGGG
"AGAG"	33	AGAAGAAGAAGA
"AGAG"	50	AGAGAGAGAGAG
"AGAG"	66	GAGGAGGAGGAG

FIG. 1. Rate constant of a DNA chain with sequence "AGAG" and GC-content of 33%, 50%, and 66% plotted as a function of the temperature and chain length. Insets show the same results in a logplot from 300 to 350 K.

bubbles and larger domains, the creation of the first bubble is an important initiation event. In a long chain, there are more places where this initial bubble could initiate, but on the other hand, the lifetime of the bubble will be small at low temperatures since the closed bps will pull the open bps back into the stack. In other words, at low temperatures and in long DNA polymers, a sequence of rare events is needed in which bubbles are first created and then they extend or new bubbles should form before the initiating bubble collapses. Extending an existing bubble is more likely than creating a new bubble due to a cooperative effect since the closed bps at the edges of a bubble open up more easily as they are pulled by an open neighbor. Still, the chance the bubble collapses will be higher than that it will grow at low temperatures. Once an open domain is formed that is beyond a critical fraction of the chain length, the remaining closed bps will have insufficient power to pull the others back into the stack but will be pulled out themselves. This implies that the longer the chain the bigger the open domain has to be to let this happen and, therefore, the rate constant decreases exponentially as a function of chain length.[33] At the higher temperatures, the lifetime of the bubbles is much higher. This implies that there is a longer time-window in which other bps can open and assist an existing bubble. Especially when the lifetime of bubbles becomes similar in magnitude as the expectation time for new bp opening events to occur, the initial bubble formation basically induces a kind of

domino effect. The rate limiting step is then determined by the formation of the initiation bubble, and since large molecules have more sites where this could happen, the rate constant actually increases as a function of chain length.

Fig. 2 shows the relative rate constants of the "AAGG," "AGGA," and "GAAG" sequences as a function of temperature and chain length relative to the "AGAG"-chain. For short sequences, the relative rates are close to 1 for all temperatures. This is not unexpected since the number of bases in which the small polymers differ is limited. For all chains, "AGAG" has the largest rate constant when the temperature is below ∼370 K, which corresponds with the results from Ref. 33 for the chains having 50% AT. When the temperature is increased, we see that the relative rates for the long chains pass first through a minimum and then increase again. The minimum corresponds to the temperature where the sequence specific order matters the most. At the higher temperatures, the rate constants become more similar to the "AGAG" results which is logical from the thermodynamics perspective that says that potential energy differences become less important at high temperatures. Unexpectedly, however, we see that "AGGA" actually surpasses "AGAG" at about 370 K. The turnover in the relative opening rates of these DNA sequences suggests that temperature might be an effective parameter to tune molecular switches based on DNA. Temperature driven DNA switches might give an easier reversible control to molecular operations than the pH driven switches that have been suggested in the literature.[40]

Despite that mesoscopic DNA models are not yet sufficiently accurate for very reliable quantitative predictions, the qualitative trend that we predict based on relative denaturation rates is an interesting phenomenon that deserves experimental efforts to test it. However, measuring denaturation rates of dsDNA is an experimental tour-de-force. The determination of hairpin denaturation rates is technically somewhat easier and, therefore, much more experimental data are available on hairpins than on dsDNA. In Sec. V B, we will therefore investigate the adapted PBD model to study hairpin denaturation and compare it with available experimental results.

B. DNA hairpins

1. Loop size dependence

As mentioned above, τ in Eq. (3) will depend not only mainly on the loop length but also slightly on the sequence as steric hindrance and solvent interactions will be important as well. Therefore, it is not straightforward to determine which value τ should have. To investigate the dependence of our model on the value of τ, we ran several simulations with different τ values and compared this trend with experimental results in which the denaturation rate constant was measured for different hairpins having the same stem but different loop-sizes. In Ref. 38, the authors measured room temperature rate constants for a series of hairpins having the same stem (GAGTCCTGGATCCTG) and loops with only T-bps but with different lengths. Fig. 3 shows the experimental denaturation rate constants as a function of the loop-length (bottom x-scale) together with the PBD calculations using the same stem as a function of τ (top x-scale). The experiments show a huge variation in measured rate constants ranging from 10^{-8}
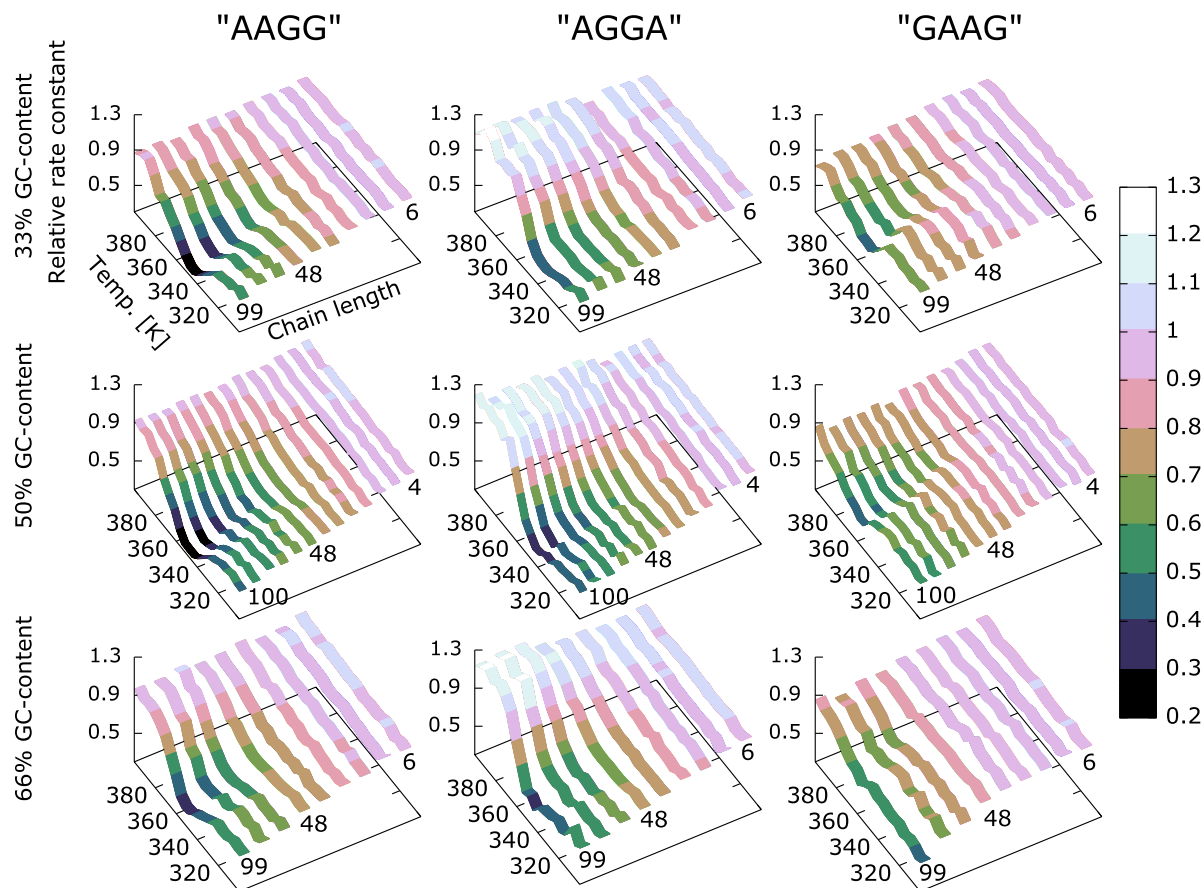
FIG. 2. Rate constant of chains "AGGA," "AAGG," and "GAAG" relative to chain "AGAG." In the first row, all chains have a GC content of 33%, in second row 50%, and third row 66%. The relative rate constant is along the z-axis, the temperature along the y-axis, and the chain length along the x-axis.

to 10^{-4} s^{-1} with an overall average tendency to increase as a function of loop-length. However, given the experimental errors, it is actually possible to fit a horizontal line striking through almost all error bar intervals (especially if we ignore the very last point having the largest error, there is only a slight increasing trend). This agrees with the statement by Bonnet et al.[39] that while the closing rate very much depends on the loop, the denaturation rate is much less affected by it. If we compare these experimental results with theoretical PBD calculations, we can observe the same trend; the rate constants increase moderately as a function of τ. A more elaborate approach has been developed in Ref. 22, where the loop is described by a Kratky-Porod model. However, this would make the calculations more expensive. Despite that we model the effect of the loop in a rather crude, simplistic manner, it seems sufficient to reproduce the main experimental feature and its orders of magnitude more efficient.

The absolute differences between the two PBD parameter sets and between the PBD and the experiments are huge. There are approximately seven orders of magnitude difference between the PBD-CaGi and the PBD-Th model. The PBD-Th results are again 8–4 orders of magnitude higher than the experimental values. When we focus on the theoretical results, we see that the PBD-CaGi parameters give a rate constant which converges very fast as a function of τ. Already for $\tau > 10$ Å, there is no noticeable change in the rate constant for increasing τ and their values correspond to that of the dsDNA model. The PBD-Th model is more flexible which

results in a longer range of τ values in which the rate constant continues to increase. Although difficult to see in the log-plot, up to $\tau = 90$ Å, the rate constant is still somewhat increasing. However, at $\tau = 19$ Å, the rate constant is already 70% of that of the dsDNA value. In the forthcoming calculations, we will stick to the values $\tau = 10$ and 19 Å for the PBD-CaGi and PBD-Th parameters, respectively. This choice is somewhat arbitrary but apparently also not so influential on the results of our calculations.
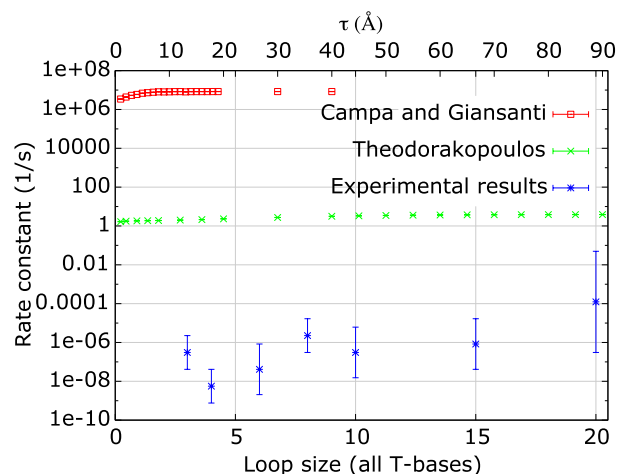

FIG. 3. Experimentally obtained rate constant (from Ref. 38) of hairpins with a 15-bps stem of 60% GC content and a T-base loop, plotted as a function of loop size. PBD model results are plotted as a function of τ.
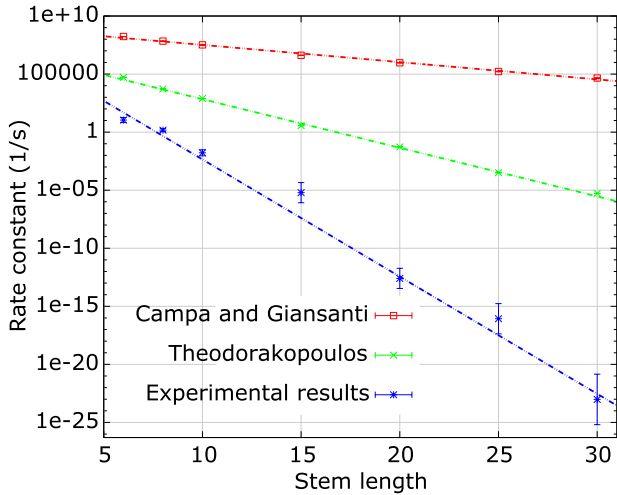
FIG. 4. Rate constant of hairpins with a stem of 50% GC-content and a loop of T4, as a function of increasing stem length, from Ref. 38. PBD model results shown together with experimental results.

2. Stem size dependence

Fig. 4 shows how the denaturation constant depends on the stem length for a DNA chain containing 50% strong GC-bps. The experimental results are taken from the same reference as before, Ref. 38. Both theoretical calculations and experiments show that the denaturation constant decreases exponentially as a function of stem length. The straight lines in Fig. 4 are exponential fits through the linear part in the log-plot. As shown in Ref. 33, the rate constant below the critical melting temperature can be expressed as

$$k \approx ab(\text{AT})^{N_{\text{AT}}}b(\text{GC})^{N_{\text{GC}}}, \quad (4)$$

with N_{AT} and N_{GC} are the number of AT- and GC-bps, respectively, and $b(\text{AT})$ and $b(\text{GC})$ are the parameters that depend mainly on temperature. Hence, $k \approx ab^N$ where a and b are the constants that depend mainly on temperature and GC-content. For 50% GC-chain, the b value follows from the homogeneous parameters as $b(50\% \text{ GC}) = \sqrt{b(\text{GC})b(\text{AT})}$ and was found to be 0.778 for the PBD-CaGi parameters at 300 K.[33] Our fit to the PBD-CaGi results in Fig. 4 resulted in $a = 10^9$ s^{-1} and $b = 0.71$.

This somewhat lower value of b is most likely due to the slightly lower temperature (298 vs 300 K) rather than due to the hairpin-effect or bp specific order. For the PBD-Th model, the exponential fit gives the values $a = 1.054 \times 10^7$ s^{-1} and $b = 0.38$ while they are $a = 5.50 \times 10^7$ s^{-1} and $b = 0.095$ for the experimental results. Hence, beside that the PBD-CaGi and PBD-Th models give an overall overestimation by orders of magnitude for all sequences, also the exponential decay as a function of chain length is too slow in comparison to experiments.

3. GC-content dependence

Reference 23 gives experimentally measured rate constants for two hairpins that differ with only one bp in the stem. The stems were 10 bps long and their sequence is shown in Table II together with their experimental and calculated rate

TABLE II. Rate constants (1/s) for two hairpins from Ref. 23. In the second sequence, one A is changed to a G with respect to the first sequence. PBD model simulations using the two parameter sets are compared with the experimental values.

Hairpin stem	PBD-CaGi	PBD-Th	Experimental value
GAAGAGGGAG	5.44×10^7	9.36×10^2	$(1.16 \pm 0.14) \times 10^{-1}$
GAAGAGGGGG	4.14×10^7	3.47×10^2	$(2.08 \pm 0.52) \times 10^{-2}$
Ratio	1.31	2.70	5.58 ± 1.55

constants. The loop was the same for both hairpins (GGGGA-GAAAGAGAGAAAGAA). The theoretical rate constants are again orders of magnitude higher than the experimental ones; however, it is interesting to compare the ratio of the rate constants of the two DNA polymers that differ only in the 9th bp. These ratios are 1.31, 2.70, and 5.58 for the PBD-CaGi, PBD-Th, and experimental rate constants, respectively. This shows again that PBD-Th is closer to the experimental value, but still a factor of 2 off regarding the relative change in rate constant when an AT-bp is changed to a GC-bp. The PBD-CaGi is another factor of 2 lower than PBD-Th. The factor 1.31 for $b(\text{AT})/b(\text{GC})$ is close to the value 1.32 at 300 K that was reported in Ref. 33 which agree well with the approximate formula in Eq. (4) for temperatures below the melting temperature.[33] Also, the experimental study of Ref. 38 reports on denaturation experiments using a series of hairpins having a loop of four T-bases with a 20-bps stem with different GC-contents. Fig. 5 shows the results together with the PBD results for the same sequences. We refer to Ref. 38 for the specific bp order of the sequences that were studied. Based on Eq. (4), we expect the following proportionality relation $k(\%\text{GC}|20 \text{ bp}) \propto q^{\%\text{GC}}$ with $q = [b(\text{GC})/b(\text{AT})]$.[19] By fitting linear lines in Fig. 5, we obtain $q = 0.052$, 3.547×10^{-9}, and 2.612×10^{-12} for the PBD-CaGi, PBD-Th, and the experimental results, respectively. Hence, based on this plot, the $b(\text{AT})/b(\text{GC})$ ratios are 1.16, 2.65, and 3.79. These values are very close to the ratios listed in Table II concerning the PBD results but also the
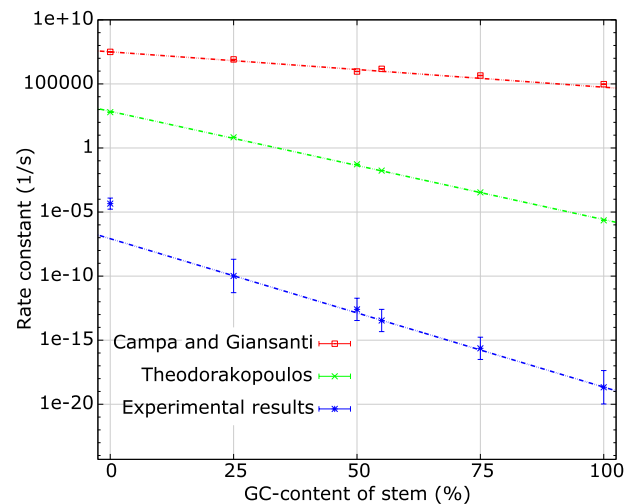

FIG. 5. Rate constant of hairpins with a 20-bps stem and a loop of T4 as a function of increasing GC-content of stem. PBD model results shown together with experimental results of Ref. 38.

experimental values agree quite well given the experimental uncertainty and the fact that these were obtained from different experiments. Combining the values for $b(AT)/b(GC)$ obtained from Fig. 5 and the values of $\sqrt{b(AT)b(GC)}$ obtained from Fig. 4, we calculated $b(GC)$ and $b(AT)$, being 0.66 and 0.76 for PBD-CaGi, 0.23 and 0.63 for PBD-Th, and 0.05 and 0.185 for the experimental results. The PBD values a little bit lower than the ones reported in Ref. 33 ($b(GC), b(AT) = 0.68, 0.90$ and 0.24, 0.64 for PBD-CAGI and PBD-Th, respectively). It is due to the fitting at shorter chain lengths and the somewhat lower temperature compared to that work.

4. Temperature dependence

In Ref. 39, Bonnet *et al.* investigated the denaturation rate constant as a function of temperature for two hairpins with identical five base-pair stem GGGAA but with one having a 21-bps loop of pure T bases and the other having a 21-bps loop of pure A bases. The experiments were performed at temperatures ranging from 285 to 310 K, and both the opening rate constant and the closing rate constant were obtained. The authors report on a slight increase in the opening rate when the loop is altered from T21 to A21, with a factor of 1.4-1.7, while the closing rate is much more affected, increasing with a factor of 2-12 when changing the loop from T21 to A21. Similar to what we noticed regarding the loop length, changing the content of the loop has a much stronger effect on the closing rate than on the denaturation rate.

As discussed in Sec. IV and shown by relation (4) is the chance that a DNA molecule denatures at low temperatures similar in terms of scaling behavior to a system in which N independent events need to occur within a short time interval; each of such events can be associated to the opening of an AT or GC bp. Naturally, this view is an oversimplification since it ignores cooperative effects or the fact that denaturation might occur in a more complex pathway in which bubbles grow, then party collapse, but then regrow again. However, these effects seem to have little influence on the overall qualitative scaling behavior of the rate constant as a function of chain length, GC-fraction, or temperature, at least for temperatures sufficiently below the melting temperature. Now, since we can associate $b(AT)$ and $b(GC)$ to the opening events of single AT- and GC-bps, it makes sense to assume Arrhenius behavior so that

$$b(AT; T) \approx c_{AT} \exp(-E_{AT}/k_B T),$$

$$b(GC; T) \approx c_{GC} \exp(-E_{GC}/k_B T),$$

$$\text{(5)}$$

where c_{AT} and c_{GC} are the proportionality constants, and E_{GC} and E_{GC} are the activation energies for the AT- and GC-bp opening, respectively. By substitution in Eq. (4), we get

$$k = a c_{AT}^{N_{AT}} c_{GC}^{N_{GC}} \exp(-[N_{AT} E_{AT} + N_{GC} E_{GC}]/k_B T). \quad \text{(6)}$$

This implies that the slope in Fig. 6 corresponds to $-(2E_{AT} + 3E_{GC})/k_B$. Now, if we assume that the proportionality constants are approximately equal in Eq. (5), $c_{AT} \approx c_{GC}$, we can actually find all unknown parameters based the previously derived values at $T = 298$ K for $b(AT)/b(GC)$ which should be equal to $\exp(-[E_{AT} - E_{GC}]/k_B 298$ K$)$.
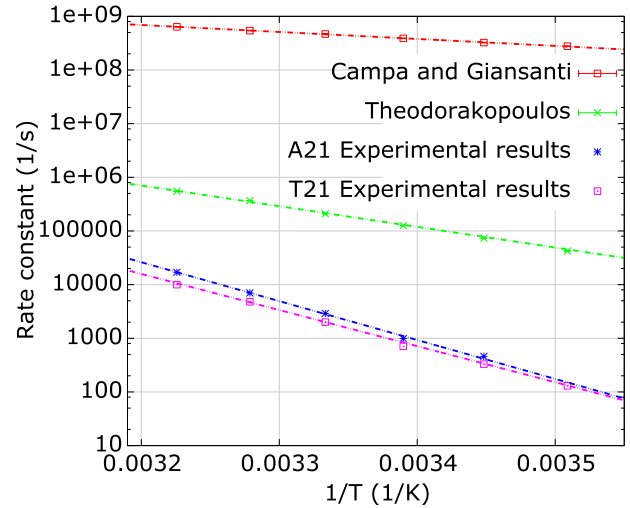

FIG. 6. Plot of rate constant for hairpins with loop of T21 and A21 as a function of the inverse temperature (from Ref. 39). PBD model results for the PBD-CaGi and PBD-Th parameter sets are presented together with the experimentally obtained results, linear fits in dashed lines.

Table III lists the different activation energies which were obtained in this way.

The top and bottom values in each block show the resulting values when we use, respectively, Table II or Fig. 5 for the ratio $b(AT)/b(GC)$ at 298 K. We see that the empirically derived activation energies correspond very well with the values of D_w and D_s, the energy barrier of the on-site Morse potentials. Hence, this suggest that in order to accurately reproduce the experimental trend, these values need to be increased till 0.29 eV for D_s and 0.25 for D_w. However, one must then take into consideration how other properties of the PBD model are affected by increasing these parameters such as denaturation curves.

5. Discussion on possible model improvements

As for all the hairpins we have investigated until now, both the PBD-CaGi parameter set and the PBD-Th parameter set yield denaturation rate constants that are orders of magnitude higher than experiments. However, we need to consider that there is an additional parameter, the friction coefficient γ, which determines the results. In this work, we have set $\gamma = 50$ ps^{-1} since it is a common value for studying biomolecules in aqueous systems. However, it is not easy to decide whether this value could be adopted for these types of mesoscopic

TABLE III. Values for E_{AT} and E_{GC} for the two parameter sets and the experimental results. All energies are given in eV.

	E_{GC}	E_{AT}	D_s	D_w
PBD-CaGi	0.0531	0.0493	0.075	0.05
	0.0527	0.0499		
PBD-Th	0.1630	0.1377	0.1655	0.1255
	0.1632	0.1375		
Expt.	0.2870	0.2525		
	0.2910	0.2465		

systems. From Stokes law, $\gamma = 6\pi\eta a$, where a is the radius of the particles and η the viscosity of the fluid. As the friction coefficient increases with the size of the particles in the system, one could argue that we need a significantly larger value for γ in these types of mesoscopic systems where each bp is described by a single coarse-grained one-dimensional particle. In high friction limit, Kramers behavior applies $\kappa \propto 1/\gamma$. Thus, the rate constant scales inversely proportional to the value of γ. Changing the friction parameter will not change the thermodynamics of the system which implies, for instance, that all theoretical denaturation curves remain the same. Since present PBD models are mostly fitted to these denaturation curves and give reasonably good results, changing the friction coefficient might seem a quick fix to our problem. However, for our particular case, we would need to enlarge the friction coefficient by several orders of magnitude to get rate constants that are in the same range as the experiments. Such a large friction coefficient is much larger than presently used in molecular simulations and is probably unphysical even for these types of mesoscopic low-dimensional models. In addition, since κ is rather insensitive to chain-length, sequence, and temperature,[33] this will not cure the incorrect scaling behavior as a function of chain length, GC content, and temperature that we observed. Still, it is a very interesting and useful question what value the friction coefficient should have for these type of models.

Improving mesoscopic models for DNA is a continuous challenge for researches in the field. A more complicated adaptation of the PBD model has been developed that includes helicity.[41] Ref. 42 extended the original PBD model to include a sequence-dependent stacking term for describing AT and GC rich regions of DNA molecules, and they derived stacking force constants for each possible stacking interaction using Monte Carlo (MC) simulations. Another adjustment is to add a narrow bulge on top of the Morse potential around the inflection point where it starts to flatten[43–45] or by replacing the Morse to get a similar overall potential.[46] If this additional potential is relatively narrow, it has the same advantage as the friction coefficient that it will not affect the thermodynamic properties a lot whereas, depending on the height of the bulge, dynamical rate constants can easily decrease by orders of magnitude.

The inclusion of this bulge has as a result that the model obtains a closing barrier which can be explained based on physical arguments. As a base moves out of stack, it gains entropy and, in addition, it will form hydrogen bonds with the solvent. In order for the base to move back into the stack, it needs to break its hydrogen bonds with the solvent and reduce its entropy again. Hence, there is a significant free energy barrier to be overcome to let this happen. In Ref. 47, the existence of this barrier was observed in free energy calculations deduced from all-atom MD simulations of DNA, and it is believable that including such a barrier in the Morse potential will result in a more realistic PBD model, but likely not sufficient. The bulge height of the potential introduced in Ref. 44 was 0.33 and 0.35 eV for AT and GC-bps, respectively, and 0.21 and 0.32 eV in Ref. 45. Reference 46 only considers pure AT-chains, using a potential with a bulge height of 0.31 eV. These values are close or even higher than the values 0.29 and 0.25 eV which we presented in Table III. Still, Ref. 46 found that the lifetime

often open bps was about 30 times too low and the lifetime of closed bps was about three orders of magnitude too low compared to experiments.[48] The disagreement of several orders of magnitude suggests that including the reclosing barrier in the PBD model is also not sufficient to describe dynamics of DNA correctly. Adding a bulge to the on-site potential also implies that base pairs that open up tend to stay open. This enhances the denaturation rate. In future work, we plan to investigate whether modifications of the Morse potential could lead to more reliable results.

VI. CONCLUSION

By exploring the nature of dsDNA with the PBD model as a function of chain length, sequence order and temperature, we found, contrary to Ref. 33, that an evenly distribution of the bps does not always have the highest denaturation rate. Especially for large sequences and at high temperatures, the dynamics change, and chains having large blocks of AT bps at the ends of the chain show favorable opening rates. We also studied a small adaptation of the PBD model in order to study denaturation of hairpins of which experimental data are available. The two parameter sets by Campa and Giansanti[7,8] and by Theodorakopoulos[9] both show denaturation rates that are orders of magnitude higher than experimental values. We also examined the denaturation rate constant as a function of loop-length, stem-length, GC-content, and temperature in the temperature range below the critical melting temperature. The PBD models give qualitatively the correct behavior regarding denaturation rates being relatively insensitive to the loop-length, exponentially decreasing as a function of stem-length, GC-content, and inverse temperature.

Our results suggests that in order for the PBD model to describe dynamics correctly, fundamental changes have to be made to the model. Further and more extensive comparison to experimental data is needed to create improved mesoscopic DNA models. This implies that besides theoretical development, also experimental progress regarding rate constant measurements is highly desired. The present experimental accuracy is presently insufficient to distinguish more subtle effect like sequence order. Our findings significate the importance of investigating dynamical quantities like the rate constant instead of only using equilibrium data for fitting model parameters on the way towards a new approach for developing more accurate mesoscopic models for DNA. The development of such a model is a prerequisite for the modeling of more complex processes like replication and transcription.

[1]International Human Genome Sequencing Consortium, "Finishing the euchromatic sequence of the human genome," Nature **431**(7011), 931–945 (2004).
[2]B. Alexandrov, N. K. Voulgarakis, K. Ø. Rasmussen, A. Usheva, and A. R. Bishop, "Pre-melting dynamics of DNA and its relation to specific functions," J. Phys. Condens. Matter **21**(3), 034107 (2009).

[3]Y. Lubelsky, J. A. Prinz, L. DeNapoli, Y. Li, J. A. Belsky, and D. M. MacAlpine, "DNA replication and transcription programs respond to the same chromatin cues," Genome Res. **24**(7), 1102–1114 (2014).

[4]T. S. van Erp, "Reaction rate calculation by parallel path swapping," Phys. Rev. Lett. **98**(26), 268301 (2007).

[5]M. Peyrard and A. R. Bishop, "Statistical mechanics of a nonlinear model for DNA denaturation," Phys. Rev. Lett. **62**, 2755–2758 (1989).

[6]T. Dauxois, M. Peyrard, and A. R. Bishop, "Entropy-driven DNA denaturation," Phys. Rev. E **47**, R44–R47 (1993).

[7]A. Campa and A. Giansanti, "Experimental tests of the peyrard-bishop model applied to the melting of very short DNA chains," Phys. Rev. E **58**, 3585–3588 (1998).

[8]A. Campa and A. Giansanti, "Melting of DNA oligomers: Dynamical models and comparison with experimental results," J. Biol. Phys. **24**(2-4), 141–155 (1999).

[9]N. Theodorakopoulos, "Melting of genomic DNA: Predictive modeling by nonlinear lattice dynamics," Phys. Rev. E **82**, 021905 (2010).

[10]R. B. Inman and R. L. Baldwin, "Helix-random coil transitions in DNA homopolymer pairs," J. Mol. Biol. **8**(4), 452–469 (1964).

[11]T. Dauxois, M. Peyrard, and A. R. Bishop, "Dynamics and thermodynamics of a nonlinear model for DNA denaturation," Phys. Rev. E **47**, 684–695 (1993).

[12]T. S. van Erp, S. Cuesta-López, and M. Peyrard, "Bubbles and denaturation in DNA," Eur. Phys. J. E **20**(4), 421–434 (2006).

[13]D. Poland and H. A. Scheraga, "Phase transitions in 1 dimension and helix-coil transition in polyamino acids," J. Chem. Phys. **45**, 1456 (1966).

[14]D. J. Proctor, H. Ma, E. Kierzek, R. Kierzek, M. Gruebele, and P. C. Bevilacqua, "Folding thermodynamics and kinetics of ynmg RNA hairpins: Specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate," Biochemistry **43**(44), 14004–14014 (2004).

[15]E. Zazopoulos, E. Lalli, D. M. Stocco, and P. Sassone-Corsi, Nature (London) **390**, 311–315 (1997).

[16]M. Menger, F. Eckstein, and D. Porschke, "Dynamics of the RNA hairpin GNRA tetraloop," Biochemistry **39**(15), 4500–4507 (2000).

[17]J. Jung and A. Van Orden, "Folding and unfolding kinetics of DNA hairpins in flowing solution by multiparameter fluorescence correlation spectroscopy," J. Phys. Chem. B **109**(8), 3648–3657 (2005).

[18]S. J. Froelich-Ammon, K. C. Gale, and N. Osheroff, J. Biol. Chem. **269**, 7719–7725 (1994).

[19]M. Takinoue and A. Suyama, "Hairpin-DNA memory using molecular addressing," Small **2**(11), 1244–1247 (2006).

[20]D. Liu and S. Balasubramanian, "A proton-fuelled DNA nanomachine," Angew. Chem., Int. Ed. **42**(46), 5734–5736 (2003).

[21]Y. Chen, S.-H. Lee, and C. Mao, "A DNA nanomachine based on a duplex-triplex transition," Angew. Chem., Int. Ed. **43**(40), 5335–5338 (2004).

[22]J. Errami, M. Peyrard, and N. Theodorakopoulos, "Modeling DNA beacons at the mesoscopic scale," Eur. Phys. J. E **23**(4), 397–411 (2007).

[23]J. Hanne, G. Zocchi, N. K. Voulgarakis, A. R. Bishop, and K. Ø. Rasmussen, "Opening rates of DNA hairpins: Experiment and model," Phys. Rev. E **76**, 011909 (2007).

[24]H. Eyring, "The activated complex in chemical reactions," J. Chem. Phys. **3**(2), 107–115 (1935).

[25]E. Wigner, "The transition state method," Trans. Faraday Soc. **34**, 29–41 (1938).

[26]J. Keck, "Statistical investigation of dissociation cross-sections for diatoms," Discuss. Faraday Soc. **33**, 173–182 (1962).

[27]D. Frenkel and B. Smit, *Understanding Molecular Simulation; From Algorithms to Applications*, 2nd ed. (Academic Press, San Diego, 2002).

[28]T. S. van Erp and P. G. Bolhuis, "Elaborating transition interface sampling methods," J. Comput. Phys. **205**, 157–181 (2005).

[29]C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, "Transition path sampling and the calculation of rate constants," J. Chem. Phys. **108**(5), 1964 (1998).

[30]T. S. van Erp, D. Moroni, and P. G. Bolhuis, "A novel path sampling method for the sampling of rate constants," J. Chem. Phys. **118**, 7762–7774 (2003).

[31]R. J. Allen, P. B. Warren, and P. R. ten Wolde, "Sampling rare switching events in biochemical networks," Phys. Rev. Lett. **94**, 018104 (2005).

[32]T. S. Van Erp, *Dynamical Rare Event Simulation Techniques for Equilibrium and Nonequilibrium Systems* (John Wiley and Sons, Inc, Hoboken, NJ, USA, 2012), Chap. 2, pp. 27–60.

[33]T. S. van Erp and M. Peyrard, "The dynamics of the DNA denaturation transition," Europhys. Lett. **98**(4), 48004 (2012).

[34]J. B. Anderson, *Predicting Rare Events in Molecular Dynamics* (John Wiley and Sons, Inc, Hoboken, NJ, USA, 2007), pp. 381–431.

[35]T. S. van Erp, "Efficiency analysis of reaction rate calculation methods using analytical models I: The two-dimensional sharp barrier," J. Chem. Phys. **125**(17), 174106 (2006).

[36]T. S. van Erp, S. Cuesta-López, J.-G. Hagmann, and M. Peyrard, "Can one predict DNA transcription start sites by studying bubbles?," Phys. Rev. Lett. **95**, 218104 (2005).

[37]R. W. Pastor, B. R. Brooks, and A. Szabo, "An analysis of the accuracy of langevin and molecular-dynamics algorithms," Mol. Phys. **65**(6), 1409–1419 (1988).

[38]M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, "Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins," Proc. Natl. Acad. Sci. U.S.A. **103**(16), 6190–6195 (2006).

[39]G. Bonnet, O. Krichevsky, and A. Libchaber, "Kinetics of conformational fluctuations in DNA hairpin-loops," Proc. Natl. Acad. Sci. U.S.A. **95**(15), 8602–8606 (1998).

[40]A. K. Jissy and A. Datta, "Designing molecular switches based on DNA-base mispairing," J. Phys. Chem. B **114**(46), 15311–15318 (2010).

[41]M. Barbi, S. Cocco, and M. Peyrard, "Helicoidal model for DNA opening," Phys. Lett. A **253**, 358 (1999).

[42]B. S. Alexandrov, V. Gelev, Y. Monisova, L. B. Alexandrov, A. R. Bishop, K. Ø. Rasmussen, and A. Usheva, Nucleic Acids Res. **37**(7), 2405–2410 (2009).

[43]G. Weber, "Sharp DNA denaturation due to solvent interaction," Europhys. Lett. **73**(5), 806 (2006).

[44]M. Peyrard, S. Cuesta-López, and D. Angelov, "Experimental and theoretical studies of sequence effects on the fluctuation and melting of short DNA molecules," J. Phys. Condens. Matter **21**(3), 034103 (2009).

[45]A. E. Bergues-Pupo, J. M. Bergues, and F. Falo, "Modeling the interaction of DNA with alternating fields," Phys. Rev. E **87**, 022703 (2013).

[46]M. Peyrard and G. James, "Intrinsic localized modes in nonlinear models inspired by DNA," NOLTA **3**(1), 27–51 (2012).

[47]E. Giudice, P. Várnai, and R. Lavery, "Base pair opening within b-DNA: Free energy pathways for GC and AT pairs from umbrella sampling simulations," Nucleic Acids Res. **31**(5), 1434–1443 (2003).

[48]M. Guéron, M. Kochoyan, and J.-L. Leroy, "A single mode of DNA base-pair opening drives imino proton exchange," Nature **328**, 89–92 (1987).