Physical Biology

OFFICIAL JOURNAL OF
Sociedad de Biofísicos Latino Americanos

PAPER

Improving the mesoscopic modeling of DNA denaturation dynamics

To cite this article: Marta Marty-Roda *et al* 2018 *Phys. Biol.* **15** 066001

IOP Publishing Phys. Biol. 15 (2018) 066001 https://doi.org/10.1088/1478-3975/aac61c

Physical Biology

PAPER

Improving the mesoscopic modeling of DNA denaturation dynamics

RECEIVED
3 January 2018

REVISED
26 April 2018

ACCEPTED FOR PUBLICATION
18 May 2018

PUBLISHED
20 June 2018

Marta Marty-Roda[1,2], Oda Dahlen[3], Titus S van Erp[3,4] and Santiago Cuesta-López[1,2,5]

[1] International Research Center in Critical Raw Materials-ICCRAM, University of Burgos, Plaza Misael Bañuelos s/n, 09001 Burgos, Spain
[2] Advanced Materials, Nuclear Technology and Applied Bio/Nanotechnology, Consolidated Research Unit UIC-154, Castilla y Leon, Spain
[3] Department of Chemistry Norwegian University of Science and Technology (NTNU) Høgskoleringen 5, 7491 Trondheim, Norway
[4] Center for Molecular Modeling (CMM), Ghent University, Technologiepark 903, B9000 Ghent, Belgium
[5] ICAMCyL Foundation, International Center for Advanced Materials and Raw Materials of Castilla y Leon, 24492 Cubillos del Sil (León), Spain

E-mail: titus.van.erp@ntnu.no **and** scuesta@icamcyl.com

Keywords: biomolecular dynamics, DNA, hairpins, modeling, PBD model, rate constants

Abstract

Although previously developed mesoscopic DNA models have successfully reproduced thermodynamic denaturation data, recent studies show that these overestimate the rate of denaturation by orders of magnitude. Using adapted Peyrard–Bishop–Dauxois (PBD) models, we have calculated the denaturation rates of several DNA hairpins and made comparison with experimental data. We show that the addition of a barrier at the onsite potential of the PBD model gives a more accurate description of the unzipping dynamics of short DNA sequences. The new models provide a refined theoretical insight on the dynamical mechanisms of unzipping which can have implications for the understanding of transcription and replication processes. Still, this class of adapted PBD models seems to have a fundamental limitation which implies that it is not possible to get agreement with available experimental results on the dynamics of long DNA sequences and at the same time maintain the good agreement regarding its thermodynamics. The reason for this is that the denaturation rate of long DNA chains is not dramatically lowered by the additional barrier—as the base-pairs that open are more likely to remain open, facilitating the opening of the full DNA molecule. Some care has to be taken, since experimental techniques suitable to the study of denaturation rates of long sequences seem not to agree with other experimental data on short DNA sequences. Further research, both theoretical and experimental, is therefore needed to resolve these inconsistencies—which will be a starting point for new minimalistic models that are able to describe both thermodynamics and dynamics at a predictive level.

1. Introduction

Understanding the physics of deoxyribonucleic acid (DNA) is a challenge since it is a high-dimensional system with a complex structure. All-atom simulations are not only computationally intensive, but also difficult to analyze. Simple models of DNA can, therefore, help to identify the essential features underlying its remarkable properties [1]. Despite the fact that a half century has passed since the discovery of the structure of the DNA double helix [2], major questions still remain regarding its thermodynamic behavior and stability. Knowledge of DNA structure alone is generally not sufficient to allow understanding of the biological processes [3, 4]. In addition to the structural information, it is necessary to know how the structure affects the equilibrium properties and the dynamics of the DNA molecule. One of the phenomena that can be studied is the local opening of DNA, which is an intriguing phenomenon from a statistical-physics point of view, but is also essential for its biological function. For instance, the unwinding of the double helix and the formation of specific DNA bubbles are prerequisites for both DNA replication and gene transcription [5].

In addition, understanding how base-pairs separate under thermal fluctuations, i.e. thermal denaturation of DNA, is of importance in many biotechnology applications, including the design of probes for primers, in polymerase chain reaction, or molecular beacons (DNA hairpins). Besides, DNA denaturation is becoming important in nanotechnology [6–9] as DNA

is now used for its self-assembly properties to create nano-devices or to design molecular memories.

In the last few decades, significant progress has been made in the development of Ising-type models [10] which describe the state of a base pair by either 0 or 1 depending on whether it is closed or open, respectively. Examples are nearest-neighbor models [11], Poland–Scheraga models [12], and Benham's model [13, 14]. Other integer-type models are the zipper models [15, 16] that assume that the DNA strand can only open from one side (neglecting the effect of bubbles). In these models, the state of the molecule is determined by a single integer variable: the number of open base-pairs. Models with a more continuous character are the bead-spring rotating strands [17] and the Peyrard–Bishop–Dauxois (PBD) model [18], which is used in the article [18]. In the latter, each base-pair is described by a single variable describing the stretch between bases. It has been shown that PBD models can reproduce the equilibrium and thermodynamic properties of DNA reasonably well [19, 21].

For a long time, the study of DNA dynamics has been given very little attention. One reason for this is the experimental difficulty of measuring kinetic data such as rates of opening or unwinding. As a result, there has been a lack of accurate experimental data about DNA dynamics to which theoretical models can be compared. Another reason is that most relevant dynamical processes, which can be compared with experiments, require very long timescales. Even if highly coarse-grained models, like PBD, can straightforwardly be used in molecular dynamics and are considerably faster than all-atom models, the accessible timescale of the simulations is generally not long enough to observe the specific biological transitions such as DNA denaturation. As a result, theoretical studies trying to analyze DNA dynamics have focused on biological phenomena with very short life-times, such as the opening of individual base-pairs [21] instead of the opening of large DNA bubbles or full molecules.

However, utilizing the fact that the PBD model has only first-neighbor interactions, the rate constants of DNA denaturation can be computed very accurately using the approach of [6, 22, 23] that is used in this work and further discussed in section 3. The same algorithmic technique has been used for the computation of denaturation rates of double stranded DNA [6, 22] and DNA hairpins [23] based on the standard PBD model. Since the latter process is more easily accessible with experimental techniques, there is significantly more experimental data available on the denaturation rate of DNA hairpins compared to double stranded DNA.

DNA hairpins are made of short single strands of DNA with terminal regions consisting of complementary base sequences. As a result, the two end-regions can self-assemble in a short DNA double helix, called the stem, while the remaining central part of the strand forms a loop. DNA hairpins have an important biological function, since they take part in many RNA configurations and are involved in key biological processes like the regulation of gene expression, DNA recombination, and regulation of transcription by binding proteins [24–26].

Several experimental studies have increased our knowledge about the kinetics of hairpin transitions by measuring the characteristic opening and closing rates [27–35]. Regarding theoretical studies, Hanne *et al* [28] compared experimental single-molecule measurements of the DNA hairpin opening rate under mechanical tension with theoretical modeling using the PBD model. Denaturation rates of short hairpins were computed in a qualitative manner using an artificial Monte Carlo type of dynamics. This approach provided trends on the behavior of relative rates as function of unzipping force and other parameters like temperature and sequence content. Dahlen and van Erp [23] were the first to report quantitative computations of hairpin denaturation rates using the PBD-class specific reactive flux implementation described in section 3. Their results show that the PBD model, despite giving reasonable results regarding thermodynamic properties, overestimates rates of unzipping by many orders of magnitude.

In this work, we will therefore examine the dynamical properties regarding hairpin denaturation using a recent adaptation of the PBD model [21] which has shown to give better agreement regarding base-pair lifetimes. In addition to the model itself, we will examine some variations of the model to explore the effect of certain model parameters on the behavior of denaturation rates as function of temperature and chain length.

2. Adapted PBD model

The PBD model describes the double stranded DNA molecule as a connected chain of 1D particles and condenses all the atomic coordinates of a base-pair into a single number y, which describes the stretching of the bonds between the two bases [3, 18]. Despite its simplicity, the model successfully describes the sharp melting transition of long DNA homopolymers [18], provides reasonable statistics on bubbles in DNA sequences [36], and on the persistence length and DNA flexibility [37]. However, the average time during which a base-pair stays closed between two opening fluctuations, and the time during which it stays open after such a fluctuation are orders of magnitude too small [21] in comparison to experimental values inferred from proton–deuterium exchange experiments [38].

One way to increase the life-times of the open state and the closed state of individual base-pairs is to add a barrier to the onsite potential. Several studies have argued that such an adaptation gives a better physical description of DNA and have proposed replacing

the original onsite Morse potential in the PBD model with a potential that has a true barrier in between the open and the closed state [21, 39, 40]. The suggested adaptations [21, 39, 40] of the potential function differ with respect to their mathematical formulations, but are rather similar in nature. Besides making it more difficult for the base-pair to open, the modification also introduces an obstacle to the closing process. Whereas, in the original PBD model, this transition is barrier-free, in the new models an energy threshold needs to be overcome. This closing barrier is a result of the hydrogen bonds that the bases make with the solvent, that need to be broken to make the transition [39, 41]. Besides enthalpic contributions, the closing barrier also reflects the loss of degrees of freedom when the base-pair gets packed into the helical structure [21], which has been quantified in all-atom free energy calculations [42]. As mesoscopic models aim to include entropy by effective potentials of mean force, the decrease in entropy is manifested by a higher potential value in the coarse-grained description. This value goes rapidly down when bringing the bases closer together, since the loss of entropy and solvent interactions is more than compensated by the favorable energetic interaction between the nearby opposite bases.

The inclusion of the barrier provides base-pair life-times that are considerably closer to experimental values [21]. Calculations on the melting transition of the A-form of DNA using the adaptation of the PBD model [21] has also given a good agreement with calorimetry and neutron scattering experiments [43]. Yet it is an open question whether this improved model, possibly with slight variations in model parameters, is able to give a correct description of the denaturation rate of DNA polymers and hairpins. The denaturation process or the opening of the full DNA chain requires many base-pair opening transitions. The opening of a single base-pair or the simultaneous opening of a few consecutive base-pairs is generally not followed up by the opening of the full molecule. As long as the majority of base-pairs are in a closed state, it is much more likely that the open base-pairs will be pulled back into the stack. The effect of the barrier on the denaturation rate is, therefore, non-trivial and has not been studied before.

The Hamiltonian for DNA chain with N base-pairs in the adapted PBD model is the following:

$$H = \sum_{n=1}^{N} \frac{p_n^2}{2m} + W(y_n, y_{n-1}) + V(y_n), \text{ with } p_n = m\frac{\mathrm{d}y_n}{\mathrm{d}t} \tag{1}$$

where $W(y_n, y_{n-1})$ for $n > 1$ is an anharmonic potential which describes the interaction between adjacent bases along the DNA molecule. $V(y)$ is the Morse potential with its modification, describing the repulsive interaction due to the hydrogen bonds between the two bases of a base-pair on opposite strands

$$W(y_n, y_{n-1}) = \frac{1}{2}K(1 + \rho e^{-\kappa(y_n + y_{n-1})})(y_n - y_{n-1})^2 \tag{2}$$

$$V(y) = D(e^{-\alpha y} - 1)^2 + \theta(y)\frac{by^3}{\cosh^2[c(\alpha y - d\ln 2)]}. \tag{3}$$

Here, $\theta(\cdot)$ is the Heaviside step-function and the original PBD model is reobtained by setting $b = 0$ in the above equation. The heterogeneity of the genetic sequence is taken into account by giving different values to the parameters of the Morse potential for adenine–thymine (AT) or guanine–cytosine (GC) base-pairs. The PBD model does not, however, distinguish between A- and T- nor between G- and C-bases. The additional barrier introduces three extra parameters b, c, and d that determine the height, the width, and the position of the added barrier. Together with the original PBD parameters (D_{AT}, D_{GC}, α_{AT}, α_{GC}, K, ρ, κ), the new model is based on ten parameters in total. Note that this is still significantly fewer than are used to fit Poland–Scheraga models. For instance [44], reports 31 fitted parameters.

In the adapted PBD model, the physically acceptable ranges for values of the width and position of the added barrier are rather narrow. Moreover, since it describes configuration space that is not frequently visited, these two parameters have no dramatic effect on either dynamics or thermodynamics. In our analysis, the c and the d variables have, therefore, been fixed to the values of [21] in this work. Moreover, rather than finding a best possible fit to experimental results using a least-square approach, in this work we rather focus on the different characteristics of the various models and investigate whether such a small modification of the PBD model actually allows the accurate description of both dynamics *and* thermodynamics.

The sets of parameters that we studied are listed in table 1. Among these are the parameter set of Campa and Giansanti [19, 22] based on the standard PBD model, the original parameters of [21], and three adaptations of this set. The last three models were studied to detect some trends in the denaturation rate when the shape of the added barrier and the Morse potential is varied, and were not a result of intensive parameter fitting. Figure 1 depicts onsite potentials used in this work.

As in [23, 28], the modeling of the DNA hairpin within the PBD framework describes the stem as a double stranded DNA molecule except for the last base-pair before the loop. This terminal base-pair is confined such that its separation y cannot extend a limiting value mimicking the stretch of the loop. Specifically, we applied the confining potential of [23] with a maximum separation for the terminal base-pair of 10 Å. In other words, the loop is not explicitly described, but only defines the maximum separation of the terminal base-pair. It has been shown that denaturation curves are not very sensitive to the maximum extension of the last base-pair [23]. In fact, the denaturation curves closely resemble that of double stranded DNA, i.e. without confining the terminal base-pair [23].

Table 1. Potential parameters used in the model.

Case	D(eV)	$\alpha(\text{Å}^{-1})$	b(eV)	$c(\text{Å}^{-1})$	d	K(eV Å^{-2})	ρ	$\kappa(\text{Å}^{-1})$
mI: Morse, parameters [19]	AT → 0.05000	4.2	0	\	\	0.025	2	0.35
	GC → 0.07500	6.9	0	\	\			
mII: Morse + barrier [21]	AT → 0.09075	3.0	4	0.74	0.2	0.004	25	0.8
	GC → 0.09900	3.4	6	0.74	0.2			
mIII: barrier 1.5 × [21]	AT → 0.09075	3.0	6	0.74	0.2	0.004	25	0.8
	GC → 0.09900	3.4	9	0.74	0.2			
mIV: barrier 2 × [21]	AT → 0.09075	3.0	8	0.74	0.2	0.004	25	0.8
	GC → 0.09900	3.4	12	0.74	0.2			
mV: [21] with deep Morse	AT → 0.1255	3.0	4	0.74	0.2	0.004	25	0.8
	GC → 0.1455	3.4	6	0.74	0.2			

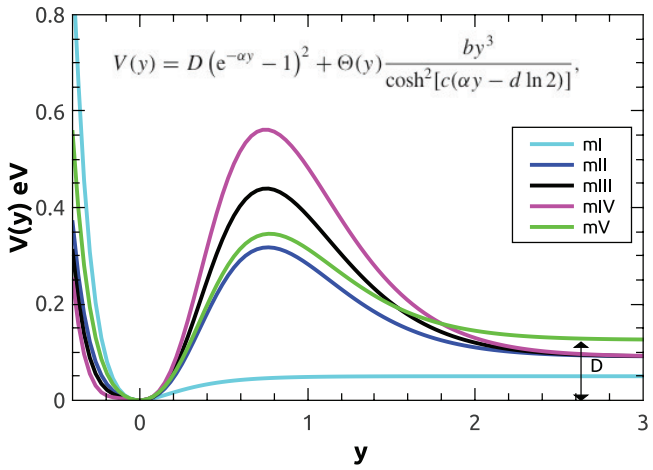


Figure 1. Morse potential used in this work for the models listed in table 1. The potentials are drawn for an homogeneous AT sequence.

3. Rate constant calculation

The reactive flux (RF) approach [45, 46] is a way to compute rate constants of processes that occur at a time-scale that is beyond the accessible time-scale of MD. The method consists of two independent calculations—a free energy calculation and the calculation of the transmission coefficient. The free energy calculation reveals the free energy profile $F(\lambda)$ as function of a well chosen reaction coordinate λ. In general, the reaction coordinate can be any non-linear function of the particles' coordinates, and its value reflects a measure of the progress of the reaction. Hence, in the PBD model, where coordinates are projected on the base-pair separations y_n:

$$\lambda = \lambda(y^N) \text{ with } y^N = \{y_1, y_2, \ldots y_N\}. \quad (4)$$

A low value of the reaction coordinate ($\lambda(y^N) < \lambda_A$) corresponds to reactant state, and a high value ($\lambda(y^N) > \lambda_B$) corresponds to the product state. In between, the transition state surface corresponds to the configurations y^N with a single value of the reaction coordinate $\lambda(y^N) = \lambda_{\text{TST}}$ that is a local maximum of the free energy profile.

The reaction constant within the reactive flux approach can be written as [45]

$$k_{AB} = \varkappa\sqrt{\frac{k_B T}{2\pi m}}\frac{e^{-\beta F(\lambda^*)}}{\int_{-\infty}^{\lambda^*} e^{-\beta F(\lambda)}\mathrm{d}\lambda} \quad (5)$$

where m is the mass that corresponds the motion of the reaction coordinate and $\beta = 1/(k_B T)$ with k_B the Boltzmann constant and T the temperature. $\varkappa$ is the transmission coefficient, which is a number between 0 and 1. The transition state theory (TST) approximation is obtained for $\varkappa = 1$ and $\lambda^* = \lambda_{\text{TST}}$.

TST assumes that each time that the transition state surface is crossed heading towards the product state, it will always reach that state without recrossing the transition state surface another time. The calculation of the transmission coefficient $\varkappa$ (which is explained below) corrects this generally invalid assumption (for the PBD model, the final rate constant is typically lower than than the uncorrected TST expression by a factor of 40 [6]). The reactive flux method is, therefore, exact and independent of the choice of reaction coordinate or the exact location of the surface λ^*. This is generally set close but not necessarily equal to λ_{TST}.

The free energy $F(\lambda')$ is related to the probability density ϱ by

$$F(\lambda') = -k_B T \ln\left[\varrho(\lambda')/\varrho_0\right], \quad (6)$$

where ϱ_0 is an arbitrary reference density to make the argument of the logarithm dimensionless. The probability density follows from an ensemble average

$$\begin{aligned}\varrho(\lambda') &= \left\langle \delta\left(\lambda(y^N) - \lambda'\right)\right\rangle \\ &= \frac{\int \mathrm{d}y^N\, \delta\left(\lambda(y^N) - \lambda'\right) \exp\left(-\beta\left[\sum_{n=1}^{N} W(y_n, y_{n-1}) + V(y_n)\right]\right)}{\int \mathrm{d}y^N \exp\left(-\beta\left[\sum_{n=1}^{N} W(y_n, y_{n-1}) + V(y_n)\right]\right)}\end{aligned} \tag{7}$$

where $\delta(\cdot)$ is the Dirac delta function.

To simplify equation (5) we can introduce the conditional probability density

$$\begin{aligned}\varrho(\lambda^*|\lambda \leqslant \lambda^*) &= \frac{\left\langle \delta\left(\lambda(y^N) - \lambda^*\right)\right\rangle}{\left\langle \theta\left(\lambda(y^N) - \lambda^*\right)\right\rangle} = \frac{e^{-\beta F(\lambda^*)}}{\int_{-\infty}^{\lambda^*} e^{-\beta F(\lambda)}\mathrm{d}\lambda} \\ &= \frac{\int \mathrm{d}y^N\, \delta\left(\lambda(y^N) - \lambda^*\right) \exp\left(-\beta\left[\sum_{n=1}^{N} W(y_n, y_{n-1}) + V(y_n)\right]\right)}{\int \mathrm{d}y^N\, \theta\left(\lambda(y^N) - \lambda^*\right) \exp\left(-\beta\left[\sum_{n=1}^{N} W(y_n, y_{n-1}) + V(y_n)\right]\right)}\end{aligned} \tag{8}$$

which can be used to substitute the exponential expressions in equation (5). The term after the first equals sign is generally the starting point for most common numerical approaches based on molecular dynamics or Monte Carlo. If computational time is not an issue, it can simply be determined by running either of those two simulations and just counting how often the reaction coordinate $\lambda(y^N)$ takes a value within a small interval $[\lambda^* : \lambda^* + \mathrm{d}\lambda]$. This number must then be divided by the number of counts that the reaction coordinate was less than or equal to λ^* and the width of the interval $\mathrm{d}\lambda$. However, since $[\lambda^* : \lambda^* + \mathrm{d}\lambda]$ interval will be rarely visited or possibly not visited at all during the whole simulation, this small probability must be determined using importance sampling methods like umbrella sampling or thermodynamic integration [45]. Despite being orders of magnitude more efficient than brute-force simulations, these approaches are often still computationally intensive as they require the convergence of many different simulations that cover relatively small parts of configuration space.

The alternative is to use the last expression of equation (8), but that would require numerical solution of the high-dimensional integrals, which is generally unfeasible. The number of grid points that need to be evaluated increases exponentially with N and the inverse integration step. However, as shown in [36], the number of required evaluations can be reduced dramatically using an iterative approach whenever the integrands can be written as product of $N-1$ 2D functions a_i which only depend on the variables y_i and y_{i-1}—in other words, whenever the integrands are of the form $a_N(y_N, y_{N-1}) \times a_{N-1}(y_{N-1}, y_{N-2}) \times a_{N-2}(y_{N-2}, y_{N-3}) \ldots \times a_3(y_3, y_2) \times a_2(y_2, y_1)$.

It is henceforth important to chose a reaction coordinate that both can be used to measure the progress of the reaction and will provide favorable integral expressions. The choice made in [22] and that we use as well is

$$\lambda(y^N) = \min(y_1, y_2, \ldots, y_N) \tag{9}$$

or the base pair separation of the least open base pair. Using this reaction coordinate equation (8) can be written as [22]

$$\begin{aligned}&\varrho(\lambda^*|\lambda \leqslant \lambda*) \\ &= \frac{\sum_i \int \mathrm{d}y^N \delta(\lambda(y_i) - \lambda^*)\left(\prod_{j\neq i}\theta(y_j - \lambda^*)\right)\prod_n \exp\left(-\beta\left[W(y_n, y_{n-1}) + V(y_n)\right]\right)}{\int \mathrm{d}y^N (1 - \prod_k \theta(y_k - \lambda^*))\prod_n \exp\left(-\beta\left[W(y_n, y_{n-1}) + V(y_n)\right]\right)}.\end{aligned} \tag{10}$$

Both the numerator and the denominator of the above expression can be written as a sum of integrals that all have the favorable product form and can, hence, be solved using Simpson's rule and the efficient iterative integration method.

The second calculation is to determine the transmission coefficient $\varkappa$, which requires the release of many molecular dynamics trajectories with starting points picked from an equilibrium distribution at the surface $\lambda(y^N) = \lambda^*$. The description above is generic for any reactive flux approach, but there are still some variations based on equivalent mathematical formulations of $\varkappa$ [47].

It is most convenient to introduce the *unnormalized* transmission coefficient R

$$R = \varkappa\sqrt{\frac{k_B T}{2\pi m}}. \tag{11}$$

Thus, equation (5) is a simple product

$$k_{AB} = R\varrho(\lambda^*|\lambda \leqslant \lambda^*). \tag{12}$$

Now, as discussed in [47] the various expressions for R can all be written as

$$R = \frac{\left\langle \dot{\lambda}\chi(y^N, p^N)\delta\left(\lambda(y^N) - \lambda^*\right)\right\rangle}{\langle \delta\left(\lambda(y^N) - \lambda^*\right)\rangle} = \left\langle \dot{\lambda}\chi(y^N, p^N)\right\rangle_{\lambda^*} \tag{13}$$

where the subscript λ^* at the bracket implies that the ensemble average has to be taken over configurations with $\lambda(y^N) = \lambda^*$. $\dot{\lambda}$ is the time-derivative of λ or the velocity along the reaction coordinate of the phase point (y^N, p^N) that is at the surface λ^*. The function $\chi(y^N, p^N)$ usually has two possible outcomes, 0 or 1, and its value is determined by a short MD trajectory starting from the phase point (y^N, p^N). As in [31], we used the *effective positive flux* formalism, which requires going both forward and backward in time. The latter is achieved by simply reverting the momenta. The expression for χ in this formalism is

$$\chi(y^N, p^N) = \begin{cases} 1 & \text{if the MD trajectory starting from } (y^N, p^N) \text{ hits } \lambda_B \\ & \text{before } \lambda_A \text{ and if the MD trajectory starting from} \\ & (y^N, -p^N) \text{ hits } \lambda_A \text{ before } \lambda^*. \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Note that whenever $\dot{\lambda} < 0$, the backward trajectory cannot cross λ_A before crossing λ^*. Hence, χ can be assigned 0 without actually running an MD trajectory in that case. If $\dot{\lambda} > 0$, the backward trajectory is performed first and then the forward trajectory only if λ_A was reached before λ^*. Hence, if the backward movement crosses λ^*, χ is immediately assigned 0, saving the need of many additional MD steps.

The effective positive flux expression not only reduces the number of MD steps per χ-evaluation, it also reduces the number of evaluations needed. Since $\dot{\lambda}\chi$ is always greater than or equal to zero for any phase point, the effective positive flux expression also converges faster than other expressions where $\dot{\lambda}\chi$ can be both positive and negative [47]. Note that if the TST assumption is valid, equation (14) reduces to $\chi = \theta(\dot{\lambda})$ and equation (13) can be solved analytically, resulting in $R = \sqrt{k_B T/2\pi m}$ as expected from equation (11).

Another algorithmic trick [22] that optimally utilizes specific characteristics of PBD-class models allows the effective generation of uncorrelated starting points (y^N, p^N) at the λ^* surface. This trick is based on the fact that the statistical ensemble with Hamiltonian H of equation (1) that is constrained to the surface $\lambda(y^N) = \lambda^*$ is equivalent to that of a freely moving chain on a translational invariant Hamiltonian H' which relates to the true Hamiltonian H via

$$H'(y^N, p^N) = H(y^N - \lambda(y^N) + \lambda^*, p^N). \tag{15}$$

Henceforth, the generation of surface points can be done by running an unconstrained MD simulation using Hamiltonian H' and storing the generated phase point every Mth time step, where M is chosen large enough to reduce correlations. Then, each stored point is first shifted by adding a single constant to all y_i values such that its minimum is at λ^*, and then MD trajectories are generated with the correct Hamiltonian H in order to compute χ for that point.

We applied this method using a numerical integration step of $\mathrm{d}y = .05$ Å for the numerical free energy calculation. Integration boundaries were set according to [6] using a tolerance of $\epsilon = 10^{-40}$, which implies that contributions with Boltzmann weight $e^{-\beta H}$ lower than this value are neglected. The phase points on the surface were generated using Hamiltonian H' and saving every $M = 1000$th step. For each rate calculation, we generated 10^6 surface points, which were used to calculate χ based on $\lambda_A = 0$, $\lambda^* = 0.75$, and $\lambda_B = 2.5$ Å. Both the generation of points using H' and the evaluation of χ using H the dynamics were influenced by a heat-bath with the requested temperature via a Langevin thermostat using a time step of 1 fs and a base-pair mass m of 300 amu. The Langevin friction was set to 50 ps^{-1}, which is considered to give accurate dynamics for biomolecules in water [48]. For friction coefficients larger than 10 ps^{-1}, the transmission coefficient and the rate constant of the standard PBD model become inversely proportional to the friction [6]. This agrees well with Kramer's theory in the high-friction regime [49]. Hence, for a friction of 50 ps^{-1}, it seems well justified to rely on overdamped Langevin dynamics (Brownian dynamics) in which the inertial terms are ignored. Brownian dynamics could provide a more efficient way to evolve the time evolution of the system, but it is not compatible with the standard approach for calculating transmission coefficients [50]. It is for this reason that we use underdamped Langevin equations of motion, even if the friction is relatively high. For an extended discussion about the friction parameter ranges used in PBD models, we refer to [51].

4. Results and discussion

Figure 2 shows the results obtained from the methodology of [6, 22, 23] and the model of equations (1) and (2) of the calculated hairpin denaturation rate constants for a set of different sequences for which experimental values are available.

Table 2 shows the sequences and temperatures of the various data points in figure 2 together with their experimental reference. Table 1 shows the parameter sets used in equation (2). Each parameter set defines a case model, denoted mI–mV. In figure 2, only models mI and mII have been shown, respectively corresponding to the original PBD model [18] with parameters from Campa and Giansanti [19] and the model with the additional barrier of [21]. Note that the model parameters of [21] were fitted to experimental data

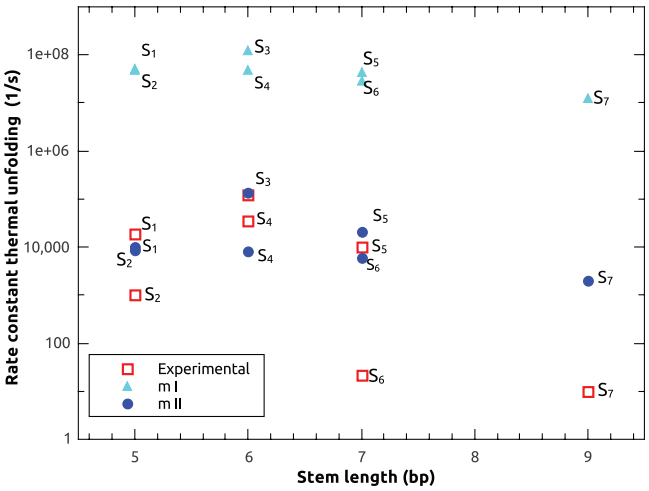


Figure 2. Thermal opening rates of hairpins from [27, 29, 30], [32–34]. Shown are the results of the original barrierless PBD model (mI) and PBD model with barrier (mII) compared with the experimental results for sequences listed in table 2. Statistical errors of the simulations are within symbol size.

on base-pair life-times and not to denaturation rates, while [19] was fitted to thermodynamic data alone.

Clearly, the adapted PBD model (mII) shows considerably better agreement with the experimental results than the original barrierless PBD model (mI). Whereas the latter is three to five orders of magnitude too high, the adapted model (mII) is either on top of the results for the short sequences (5 base-pair stem) or maximally two orders of magnitude off for the longer ones (9 base-pair). It is important to realize that the experimental data also show huge variations for sequences having the same stem length and same GC content. This could be due to the differences in loop length or sequence, or due to variations in the temperatures under which the experiments were performed. However, it is also not unthinkable that the different experimental measurement techniques perturb the dynamics and, therefore, will possibly not provide the same rates. We come back to this when we compare our results with that of another experimental study based on optical tweezers [35].

In figure 3, we examine the denaturation rate as function of temperature. Experimental results were taken from [27] by Bonnet *et al* who investigated two hairpins with identical five base-pair stem GGGAA but with loops either consisting of 21 A-bases or 21 T-bases. Note that these two systems have identical descriptions in our models, not only because it does not distinguish between A and T bases, but also because the loop itself is not explicitly described, as explained above. In figure 3(a) we use models mI and mII of table 2 which do not involve any parameter fitting. The results indicate that the original PBD model deviates from the experimental results by five and three orders of magnitude at low (ca. 280 K) and high temperatures (ca. 315 K), respectively. The model with additional barrier shows, for the whole range of temperatures, results that are within one order of magnitude of the experimental values. From the slope and the intersection

Table 2. (a) Sequences and temperatures of thermal unfolding hairpins. (b) Sequences of hairpins used in optical tweezer experiments.

(a) Thermal denaturation exp. [27–34]

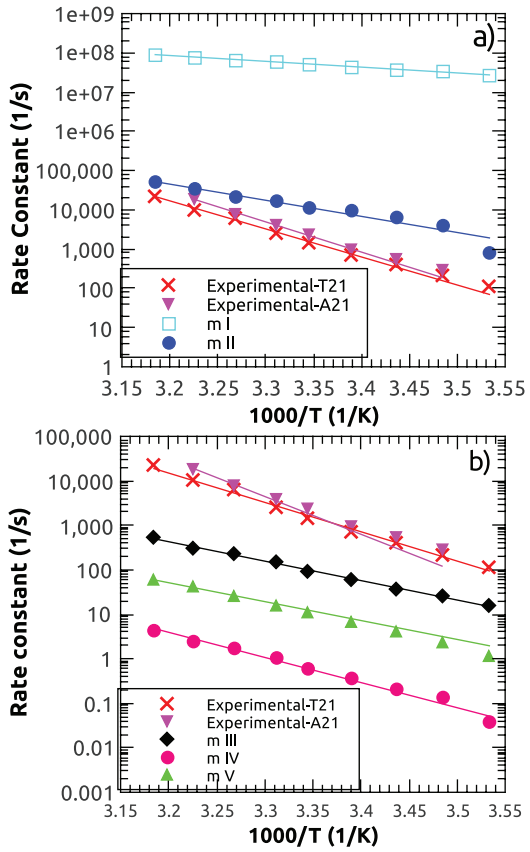
Label	Stem	Loop	Temp.(°C)	References
S_1	AAGGG	T_{21}	25	[29]
S_2	GGGAA	$T_{n=4}$;A_{30}	23	[27, 33]
S_3	AAGGAA	T_4	54	[30]
S_4	GGAAAA	T_4	23	[31]
S_5	GGGAAAA	T_8	36	[32]
S_6	GAGAAGA	A_{40}	25	[34]
S_7	GAGAAGAGA	A_{40}	25	[34]

(b) Optical tweezer exp. [35]

Length	Stem	Temp. = 23°C
6	GAGGAA	
8	GAGAGGAA	
10	GAGAGAGGAA	
15	GAGAGGAGGAAGGAA	
20	GAGAGAAGGAGAGGAAGGAA	
25	GAGAGAAGGAAGAGGGAGGAAGGAA	
30	GAGAGAAGGAAGAGAAGAGGGAGGAAGGAA	

of the straight fitted lines, we obtained the Arrhenius parameters, the activation energy E_a and the pre-exponential factor A, which are shown in figure 3(c). It shows that the mismatch between theory and experiment is largely due to the pre-exponential factor.

Overall, this shows that the model proposed in [21] considerably improves the dynamics of DNA and, especially for short DNA sequences, reproduces the experimental denaturation rate constants reasonably well. Figure 3(b) shows the same experimental results again, together with models mIII–mIV. Model mIII is nearly as good as mII with the difference that mIII gives rate constants that are lower than the experimental results, while the mII results are higher. On the logarithmic scale, the curves of the rate constants



Fitted Arrhenius parameters

Case	E_a $(\frac{\text{kJ}}{\text{mol}})$	A $(\frac{1}{\text{s}})$
Exp. T_{21}	130.61	$1.85{\cdot}10^{25}$
Exp. A_{21}	135.10	$9.85{\cdot}10^{26}$
mI	28.26	$4.63{\cdot}10^{12}$
mII	84.63	$7.07{\cdot}10^{18}$
mIII	83.63	$4.09{\cdot}10^{16}$
mIV	105.50	$1.72{\cdot}10^{18}$
mV	81.31	$2.02{\cdot}10^{15}$

Figure 3. Opening constant of hairpins with stem GGGAA and loop of T21 and A21 as function of the inverse temperature (from [27]) and comparison with our model with various sets of parameters. (a) Experimental data compared with theoretical results from the original PBD model (mI) with parameters from [19, 20] and the adapted PBD model (mII) with parameters from [21]. (b) Comparison of the same experimental data with the theoretical results of the adapted PBD model with parameter variations (models mIII–mV of table 1).

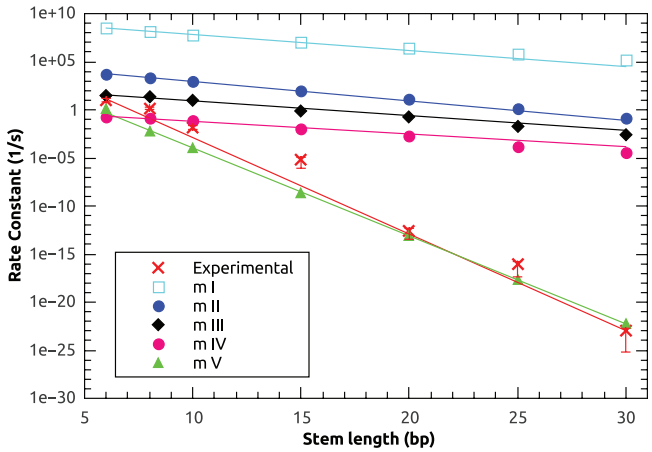


Figure 4. Rate constant of denaturation measured by unloaded optical tweezer experiments [35] and calculated by thermal denaturation simulation using models mI–mV of table 1.

as function of the inverse temperature have nearly the same slopes.

On the other hand, if we compare the rate versus stem length we see a different behavior. Figure 4 shows for various stem lengths the denaturation rate of seven experimental DNA hairpins. The stem sequences are given in table 2(b) and contain 50% GC base-pairs and a 4T loop sequence at a temperature of 23°C. The experimental denaturation rates are taken from the data of unloaded (in absence of external force) optical tweezer experiments [35], which show a steep exponential decay as function of chain length. The PBD models mI and mII produce a much weaker decay as function of chain length, which has as a consequence that the deviation between the model and the experimental results increases beyond 15 orders

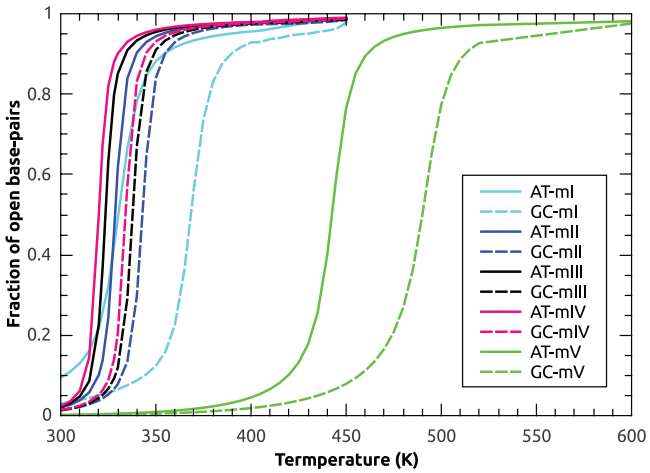


Figure 5. Fraction of open base-pairs as function of temperature for long (400 base-pairs) homopolymers calculated using the approach of [5] with models mI–mV of table 1.

of magnitude for the longest chain in this data set (30 base-pairs). The model with barrier [21] is around five orders of magnitude lower than the original PBD model for all sequences, i.e. closer to the experimental results, but it fails to reproduce the same slope in the $\ln k$ versus chain length plot. To investigate whether the shape of the barrier could be adjusted to reproduce the same experimental trend, we looked at models mIII and mIV (see table 1), in which we increase the barrier height by factors of 1.5 and 2 respectively.

Clearly, the barrier increase gives an overall shift in the logarithmic plot but hardly changes the slope. The reason for this is that this increase has two opposing effects: not only will it slow down the opening rate of individual base-pairs, it also makes it more difficult for this base-pair to reclose. The latter effect increases the denaturation rate of the whole DNA sequence, and is the reason for the decay being independent of the barrier height (parameter b in equation (2)). In order to get qualitatively correct behavior of the denaturation rate versus chain length, there is no other option than to change the relative probabilities between the open and closed states for the individual base-pairs. Equivalently, the barrier to opening should be increased without increasing the closure barrier with the same amount. To do this, we changed the depth of the Morse potential (or equivalently, as the minimum is set at zero, we increased the height of the plateau)—leading to model mV, which can quantitatively reproduce the optical tweezer experiments at all lengths.

Changing the height of the Morse potential comes with a price, as can be seen in figure 5 where we show the denaturation curves of long (400 base-pair) homopolymers, showing the typical phase transition behavior [18]. Experimentally, this transition occurs around 70 °C for pure GC sequences [52] (with variations depending on the ion concentration), but model mV gives melting transitions for AT and GC chains that are beyond the boiling temperature of water. Hence, this shows that it is impossible to get both the thermodynamics and dynamics correct at the same time using a single unified mesoscopic model. Improvements might lie in the replacement of the flat horizontal Morse plateau with a function having negative slope to describe entropic effects yet unconsidered, such as rotational degrees of freedom and the possibility for open regions to curl up. Ideally, the development of such a model should be derived from full atomistic simulations, but the computational expense for simulating long DNA sequences in explicit water is yet too large. The alternative is to use a more complex model and fit it to experimental data. This should be done on both thermodynamic and dynamical data. Unfortunately, the experimental uncertainties regarding denaturation rates are still very large, which makes it difficult to decide which characteristics the theoretical models should have. The deviation in experimental results is obvious from figure 2. The discrepancy is even larger if we compare those results with the results of the optical tweezer experiments with zero load [35] of figure 4 for the short sequences. This comparison shows that the optical tweezer experiments give results two to three orders of magnitude lower compared to the thermal denaturation experiments for short chain lengths. For stem lengths of six base-pairs, the optical tweezer experiment gives results close to those of models mIII, mIV, mV (figure 4) However, for longer stem lengths, only the model with the deep Morse potential can reproduce the optical tweezer experiments.

5. Conclusion

We have shown that a recent adaptation [21] of the PBD model with an additional barrier to the Morse potential improves the agreement of the denaturation rate for DNA hairpins in comparison with experimental results. For short sequences, this agreement is relatively good, but the magnitude of the exponential decay of the rate as function of chain length does not correspond to those measured by

optical tweezer experiments [35]. As a result, the overestimation of denaturation rates of the model in comparison to experiments increases with increasing chain length. As we show, it is possible to change the model parameters such that they reproduce the right decay, but this increases the melting temperature of the long homopolymers to unrealistic values. Therefore, a more complex model should ideally be fitted to both thermodynamic and dynamical data, especially since the latter are much more sensitive to model parameters. The data could be denaturation rates, but other dynamical properties could be also used for fitting, like thermal conductance [9]. Possibly a slightly more complex helicoidal model [53] that more explicitly included the effects of bending, torsional energies, and unwinding, would be more successful in describing both dynamics and thermodynamics at the same time. Another possibility that could resolve the mismatch between thermodynamics and dynamics is to let the model parameters be temperature dependent. However, fitting such a model, while maintaining some kind of physical justification, is far from trivial.

An essential condition for establishing better models is an increase in both quantity and quality of experimental data on dynamics. Rate constants are more difficult to measure in comparison with thermodynamic properties, and error-bars are thus typically a few order orders of magnitude. For a long time, rate constants from computer simulations were equally inaccurate, but due to increasing computer power and algorithmic innovations this has changed. Especially for models of the first-neighbor type, rate constants can be obtained with very low statistical errors using the method described in this article. Therefore, a similar improvement in experimental accuracy would be very welcome and allow for a more reliable parameter fitting of theoretical models.

More disturbing from a theoretician's point of view is that experimental rate constants based on diverse experimental techniques do not always agree, even when considering the large error-bars. This field would therefore benefit from more systematic or automated experimental approaches. Defining a few standardized reference conditions related to ion concentrations and temperature would facilitate the comparison of results between different experimental groups. High-throughput screening could eliminate human factors in e.g. sample preparation and reduced error-bars, due to an increase of repetitive identical measurements. In addition to this, it would be good to find estimates about the effect of the experimental probe on the results, and on the validity of some theoretical assumptions, like the infinite dilution limit, which neglects interactions between different DNA molecules. Finally, the design of theoretical models would be highly facilitated by an increase of experimental data on synthetically created 'simple' DNA sequences, like homogeneous A- or G-sequences of various length, alternating (AGAGAG...) sequences, and the ones reported in [6, 23]. We therefore hope that our results will inspire both experimentalists and theoreticians to study dynamical transitions in DNA.

Acknowledgments

The authors would like to thank Professor Peyrard for valuable, helpful and instructive discussions. MMR and SCL acknowledge the overall support by the Spanish Ministry of Economy, Industry and Competitiveness (BES-2013-065453, FIS2012-38827) and the University of Burgos and the Anders Jahre fund (Project 40105000) are thanked for financial support of MMR's research secondments at the NTNU Trondheim. In addition, SCL and UC-154 are grateful for the Ramon y Cajal grant (2009) and the support of Junta de Castilla y Leon (Spain). OD and TSE acknowledge the Research Council of Norway (projnr. 10422600).

ORCID iDs

Oda Dahlen https://orcid.org/0000-0002-1048-0605
Titus S van Erp https://orcid.org/0000-0001-6600-6657
Santiago Cuesta-López https://orcid.org/0000-0002-7401-3889

References

[1] Cuesta-Lopez S, Peyrard M and Graham D 2005 Model for DNA hairpin denaturation *Eur. Phys. J.* E **16** 235–46
[2] Watson J D *et al* 1953 Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid *Nature* **171** 737–8
[3] Peyrard M and Bishop A R 1989 Statistical mechanics of a nonlinear model for DNA denaturation *Phys. Rev. Lett.* **62** 2755–8
[4] Alexandrov B, Voulgarakis N K, Rasmussen K Ø, Usheva A and Bishop A R 2009 Pre-melting dynamics of DNA and its relation to specific functions *J. Phys.: Condens. Matter* **21** 034107
[5] van Erp T S, Cuesta-Lopez S and Peyrard M 2006 Bubbles and denaturation in DNA *Eur. Phys. J.* E **20** 421–34
[6] van Erp T S and Peyrard M 2012 The dynamics of the DNA denaturation transition *Europhys. Lett.* **98** 48004
[7] Markham N R and Zuker M 2005 Dinamelt web server for nucleic acid melting prediction *Nucleic Acids Res.* **33** W577–81
[8] Takinoue M and Suyama A 2006 Hairpin-DNA memory using molecular addressing *Small* **2** 1244–7
[9] Chien C C, Velizhanin K A, Dubi Y and Zwolak M 2013 Tunable thermal switching via DNA-based nano-devices *Nanotechnology* **24** 095704
[10] Ising E 1925 Beitrag zur theorie des ferromagnetismus *Z. Phys.* A **31** 253–8
[11] SantaLucia J 1998 A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics *Proc. Natl Acad. Sci. USA* **95** 1460–5
[12] Poland D and Scheraga H A 1966 Phase transitions in one dimension and the helixcoil transition in polyamino acids *J. Chem. Phys.* **45** 1456–63
[13] Benham C J 1979 An elastic model of the large-scale structure of duplex DNA *Biopolymers* **18** 609–23

[14] Fye R M and Benham C J 1999 Exact method for numerically analyzing a model of local denaturation in superhelically stressed DNA *Phys. Rev.* E **59** 3408
[15] Ivanov V, Zeng Y and Zocchi G 2004 Statistical mechanics of base stacking and pairing in DNA melting *Phys. Rev.* E **70** 051907
[16] Kittel C 1969 Phase transition of a molecular zipper *Am. J. Phys.* **37** 917–20
[17] Dasanna A K, Destainville N, Palmeri J and Manghi M 2013 Slow closure of denaturation bubbles in DNA: twist matters *Phys. Rev.* E **87** 052703
[18] Dauxois T, Peyrard M and Bishop A R 1993 Entropy-driven DNA denaturation *Phys. Rev.* E **47** R44
[19] Campa A and Giansanti A 1998 Experimental tests of the peyrard-bishop model applied to the melting of very short DNA chains *Phys. Rev.* E **58** 3585
[20] Campa A and Giansanti A 1999 Melting of DNA oligomers: dynamical models and comparison with experimental results *J. Biol. Phys.* **24** 141–55
[21] Peyrard M, Cuesta-Lopez S and Angelov D 2009 Experimental and theoretical studies of sequence effects on the fluctuation and melting of short DNA molecules *J. Phys.: Condens. Matter* **21** 034103
[22] van Erp T S 2007 Reaction rate calculation by parallel path swapping *Phys. Rev. Lett.* **98** 268301
[23] Dahlen O and van Erp T S 2015 Mesoscopic modeling of DNA denaturation rates: sequence dependence and experimental comparison *J. Chem. Phys.* **142** 235101
[24] Zazopoulos E, Lalli E, Stocco D M and Sassone-Corsi P 1997 Dna binding and transcriptional repression by dax-1 blocks steroidogenesis *Nature* **390** 311–5
[25] Trinh T Q and Sinden R 1993 The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in escherichia coli *Genetics* **134** 409 (PMID: 8325478)
[26] Proctor D J, Ma H, Kierzek E, Kierzek R, Gruebele M and Bevilacqua P C 2004 Folding thermodynamics and kinetics of ynmg rna hairpins: specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate *Biochemistry* **43** 14004–14
[27] Bonnet G, Krichevsky O and Libchaber A 1998 Kinetics of conformational fluctuations in DNA hairpin-loops *Proc. Natl Acad. Sci. USA* **95** 8602–6
[28] Hanne J, Zocchi G, Voulgarakis N K, Bishop A R and Rasmussen K Ø 2007 Opening rates of DNA hairpins: Experiment and model *Phys. Rev.* E **76** 011909
[29] Jung J and Van Orden A 2005 Folding and unfolding kinetics of DNA hairpins in flowing solution by multiparameter fluorescence correlation spectroscopy *J. Phys. Chem.* B **109** 3648–57
[30] Hilbers C W, Haasnoot C A G, De Bruin S H, Joordens J J M, Van Der Marel G A and Van Boom J H 1985 Hairpin formation in synthetic oligonucleotides *Biochimie* **67** 685–95
[31] Ansari A, Kuznetsov S V and Shen Y 2001 Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins *Proc. Natl Acad. Sci. USA* **98** 7771–6
[32] Ansari A and Kuznetsov S V 2005 Is hairpin formation in single-stranded polynucleotide diffusion-controlled? *J. Phys. Chem.* B **109** 12982–9
[33] Wallace M I, Ying L, Balasubramanian S and Klenerman D 2001 Non-Arrhenius kinetics for the loop closure of a DNA hairpin *Proc. Natl Acad. Sci. USA* **98** 5584–9
[34] Grunwell J R, Glass J L, Lacoste T D, Deniz A A, Chemla D S and Schultz P G 2001 Monitoring the conformational fluctuations of DNA hairpins using single-pair fluorescence resonance energy transfer *J. Am. Chem. Soc.* **123** 4295–303
[35] Woodside M T, Behnke-Parks W M, Larizadeh K, Travers K, Herschlag D and Block S M 2006 Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins *Proc. Natl Acad. Sci. USA* **103** 6190–5
[36] van Erp T S, Cuesta-Lopez S, Hagmann J G and Peyrard M 2005 Can one predict DNA transcription start sites by studying bubbles? *Phys. Rev. Lett.* **95** 218104
[37] Theodorakopoulos N and Peyrard M 2012 Base pair openings and temperature dependence of DNA flexibility *Phys. Rev. Lett.* **108** 078104
[38] Leroy J L, Kochoyan M, Huynhdinh T and Gueron M 1988 Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton *J. Mol. Biol.* **200** 223–38
[39] Weber G 2006 Sharp DNA denaturation due to solvent interaction *Europhys. Lett.* **73** 806
[40] Peyrard M and James G 2012 Intrinsic localized modes in nonlinear models inspired by DNA *Nonlinear Theory Appl. IEICE* **3** 27–51
[41] Drukker K, Wu G S and Schatz G C 2001 Modes simulations of DNA denaturation dynamics *J. Chem. Phys.* **114** 579–90
[42] Giudice E, Varnai P and Lavery R 2003 AT pairs from umbrella sampling simulations *Nucleic Acids Res.* **31** 1434–43
[43] Valle-Orero J, Wildes A R, Theodorakopoulos N, Cuesta-López S, Garden J L, Danilkin S and Peyrard M 2014 Thermal denaturation of a-DNA *New J. Phys.* **6** 113017
[44] Jost D and Everaers R 2009 Polynucleotide DNA melting: salt effects and predictive power *Biophys. J.* **96** 1056–67
[45] Frenkel D and Smit B 2002 *Understanding Molecular Simulation* 2nd edn (San Diego, CA: Academic)
[46] Chandler D 1987 *Introduction to Modern Statistical Mechanics* (New York: Oxford University Press)
[47] Van Erp T S 2012 Dynamical rare event simulation techniques for equilibrium and nonequilibrium systems *Adv. Chem. Phys.* **151** 27
[48] Pastor R W, Brooks B R and Szabo A 1988 An analysis of the accuracy of langevin and molecular-dynamics algorithms *Mol. Phys.* **65** 1409–19
[49] Hanggi P, Talkner P and Borkovec M 1990 Reaction-rate theory—50 years after Kramers *Rev. Mod. Phys.* **62** 251–341
[50] Menzl G, Singraber A and Dellago C 2016 S-shooting: a Bennett–Chandler-like method for the computation of rate constants from committor trajectories *Faraday Discuss.* **195** 345–64
[51] Manghi M and Destainville N 2016 Physics of base-pairing dynamics in DNA *Phys. Rep.* **631** 1–41
[52] Inman R and Baldwin R 1964 Helix-random coil transitions in DNA homopolymer pairs *J. Mol. Biol.* **8** 452–69
[53] Barbi M, Cocco S and Peyrard M 1999 Helicoidal model for DNA opening *Phys. Lett.* A **253** 358–69