

DNA melting: Intra-base-pair dynamics and a vector generalization of the Peyrard-Bishop-Dauxois model

Nikos Theodorakopoulos 

Theoretical and Physical Chemistry Institute, National Hellenic Research Foundation, Vasileos Constantinou 48, 116 35 Athens, Greece and Fachbereich Physik, Universität Konstanz, 78457 Konstanz, Germany



(Received 6 October 2025; revised 12 December 2025; accepted 24 April 2026; published 11 May 2026)

The Peyrard-Bishop-Dauxois (PBD) model of DNA denaturation, although successful in the description of melting profiles, fails to predict melting entropies, unzipping forces, and dynamical properties, e.g., hairpin dynamics. The paper presents an atomistic “toy model” of the intra-base-pair motion which suggests that the thermodynamics may be better described by a planar vector—rather than a scalar—order parameter. This leads to correct estimates of melting entropy, unzipping force, hairpin opening rates, and the equilibrium constant of open/closed base pair states during imino proton exchange.

DOI: [10.1103/7mtw-wvkx](https://doi.org/10.1103/7mtw-wvkx)

I. INTRODUCTION

Mesoscopic modeling is a key tool in the exploration of important biophysical processes which occur at or above the nanoscale. The Peyrard-Bishop-Dauxois (PBD [1]) approach to DNA denaturation (“melting”) is a representative example of such modeling. It was motivated by nonlinear lattice dynamics (“phonon softening”) and deals with the dynamics of each base pair in terms of a single scalar variable, which provides an effective measure of the spatial separation of the two bases. The combination of a local, effective intrapair interaction (modeled by a Morse potential and describing the combined effects of hydrogen bonding, stacking, and solvent) and a nonlocal, nonlinear elastic interaction between nearest neighbor pairs (accounting for the effects of base-pair stacking) leads to one of the simplest Hamiltonian models which exhibit an exact thermodynamic phase transition in one dimension. Such a transition is predicted to occur in the case of “homogeneous” DNA, i.e., a polynucleotide of infinite extent [2,3]. The heterogeneity of natural DNA leads to a rounding of the transition, typically of the multistep type [4]. Theoretical calculations with a revised set of model parameters have successfully matched observed melting profiles of long genomic chains with a rich internal structure, in solution [5] as well as in fibers [6]. It would thus seem that the model has matured beyond “proof of concept” toward practical applicability.

Nonetheless, important discrepancies persist: First, the calculated melting entropy [5] is less than half the calorimetrically measured transition entropy [7,8]. A similar discrepancy exists between model predictions [9] and experimental data [10] regarding the magnitude of the force required for unzipping the two DNA strands. Furthermore, rate constants of

DNA hairpin dynamics calculated [11] with the same set of model parameters as in Ref. [5] are overestimated by several orders of magnitude compared to experiment; this has motivated further efforts [12] to improve on the model by adding a barrier to the on-site potential; however, this feature, although it improves agreement with experiments on the dynamics of short chains, does not seem to work for long chains.

This work has been motivated by the idea that the failure of the original PBD model to describe hairpin dynamics could be related to its incomplete description of equilibrium properties, namely the entropy and unzipping force discrepancies mentioned above, and that these discrepancies must be addressed first. In the next section, I will provide an alternative view of the microscopic, intra-base-pair dynamics underlying the original PBD model. The proposed atomistic picture modifies the mesoscopic coarse graining, leading to the identification of two relevant degrees of freedom per base pair; to some extent, the two oscillators per base pair mimic the physical situation with its two or three hydrogen bonds. Equivalently, this reduction of the double helix statistical mechanics can be thought of as a planar vector PBD chain with strongly anisotropic local and nonlocal nonlinearities, and a concomitant doubling of the calculated extensive thermodynamic quantities. Section III presents the calculation of the entropy in a standard case. Section IV presents results on the unzipping force. Section V addresses the questions raised by hairpin opening rates. Concluding remarks are made in the final section and include an interpretation of the equilibrium constant between open and closed states of a base pair during NMR imino proton exchange experiments [13].

II. RETHINKING PBD

A. A toy model of the base pair plane

The model is illustrated in Fig. 1; each of the two complementary bases consists of a main part and two attachments of smaller mass held rigidly at a distance l from the center of

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

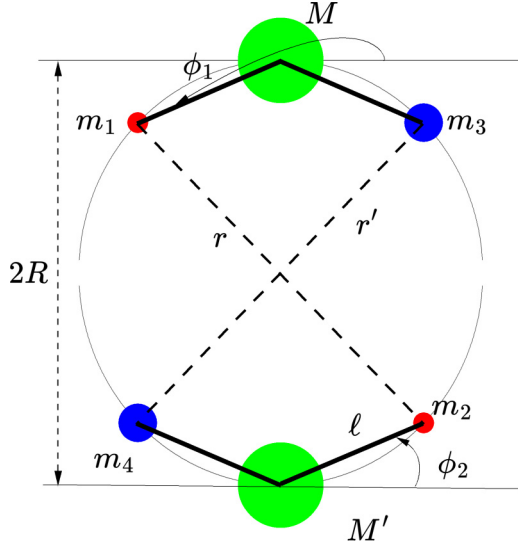


FIG. 1. A toy model of a single base plane in the DNA double helix. The large full circles represent the core of each base, of mass M and M' , respectively, and are positioned at diametrically opposite points of a circle of radius R . Attached to each base at a distance $l = 2R \sin(\pi/8)$ are two point masses, one smaller, mimicking the carbonyl and exocyclic amino groups, and another approximating endocyclic nitrogen and its surrounding unit. The angle between the bond connecting the mass m_i to its neighboring base core and the horizontal line is ϕ_i ; at equilibrium, $\phi_2^{(0)} = \phi_1^{(0)} - \pi = \pi/8$. The dashed lines represent the effective couplings between the complementary bases; they are perpendicular to each other at equilibrium. Since the mass ratios m_i/M , $m_i/M' \ll 1$ the core parts of the bases are approximately fixed in space. At equilibrium, all four mass points are on the perimeter of the circle and the distances r and r' are equal to the circle's diameter $2R$.

the main part. The direction of the bond between main part and attachment is free to vary, but is assumed to remain on the base plane. The attachments with the lowest mass m_1, m_2 are meant to mimic the carbonyl and the exocyclic amino groups, which are responsible for the hydrogen bonding between complementary bases and therefore typically have a mass of about 16 amu, far smaller than the remaining mass M of a base, which is of the order of 300 amu. The attachments with the intermediate masses m_3, m_4 are meant as approximations to endocyclic nitrogen atoms and their surrounding units. Again, it is assumed that these masses too are substantially lower than M . This means that the dynamics of the entire planar system can be well approximated by the dynamics of the four point masses m_i , ($i = 1, 2, 3, 4$), while the main parts of the bases can be thought of as being fixed in space. The dashed lines represent the effective coupling between complementary bases; they are perpendicular to each other at equilibrium.

I will use a Cartesian coordinate system with the center of the circle (Fig. 1) as the origin. The positions of the four mass points m_i can be described in terms of the direction angles ϕ_i relative to the horizontal axis. Noting that the motion of the masses 1 and 2 is decoupled from that of 3 and 4, I can write the Lagrangian of the former as

$$\mathcal{L} = \frac{\ell^2}{2} \sum_{i=1}^2 m_i \dot{\phi}_i^2 - V(r - r_0), \quad (1)$$

where

$$\left(\frac{r}{2R}\right)^2 = 1 + \left(\frac{\ell}{R}\right)^2 \sin^2\left(\frac{\phi_1 - \phi_2}{2}\right) + \frac{\ell}{R}(\sin \phi_1 - \sin \phi_2), \quad (2)$$

V is the interaction potential between the two point masses, r their separation distance, and $r_0 = 2R$ the latter's equilibrium value. As in the PBD model, I will use a Morse potential with depth and width parameters D and $1/\alpha$, respectively,

$$V(\eta) = D(1 - e^{-\alpha\eta})^2$$

and its harmonic limit

$$V_h(\eta) = \frac{1}{2}K\eta^2,$$

where $K = 2D\alpha^2$.

The equations of motion corresponding to (1) are

$$m_i \ell^2 \ddot{\phi}_i = -\frac{\partial V}{\partial \phi_i}, \quad i = 1, 2. \quad (3)$$

At equilibrium, $r = r_0 = 2R$,

$$\phi_2 = \phi_2^{(0)} = \phi_1^{(0)} - \pi = \arcsin \frac{l}{2R} = \pi/8.$$

A completely analogous set of equations describes the dynamics of the subsystem composed of masses 3 and 4.

In the limit of small oscillations around equilibrium, $\delta\phi_i = \phi_i - \phi_i^{(0)} = \delta\phi_i^{(0)} e^{i\omega t}$, the equations of motion (3) lead to two distinct eigenmodes. One eigenmode has a zero eigenvalue, and its eigenvector components satisfy the relationship

$$\delta\phi_1^{(0)} + \delta\phi_2^{(0)} = 0.$$

The other eigenmode has a nonzero oscillation frequency

$$\omega^2 = K \cos^2(\pi/8) \left(\frac{1}{m_1} + \frac{1}{m_2} \right), \quad (4)$$

and the corresponding unnormalized eigenvector satisfies

$$\frac{m_1}{m_2} \delta\phi_1^{(0)} - \delta\phi_2^{(0)} = 0.$$

It is straightforward to see that the zero frequency mode reflects the local invariance of the Hamiltonian to small, equal, and opposite rotations of the bonds ($Mm_1, M'm_2$); the latter can be seen to leave the distance $r(\phi_1, \phi_2)$ unchanged. Any nontrivial dynamics of the subsystem $\{m_1, m_2\}$ originates in the nonzero frequency eigenmode. This in turn, except for the geometrical factor, is just the dynamics of a PBD “diatomic pair” coupled by the interaction potential V .

Identical arguments hold for the subsystem $\{M', m_3, m_4\}$; the only difference being that the oscillator frequency (4) will now be lower on account of the higher masses.

B. A two-chain version of the PBD model

The toy model presented above may serve as an “atomistic foundation” of a variant of the original PBD model. In this variant, there is one more independent oscillator per base pair, originating in the motion of the subsystem $\{M', m_3, m_4\}$. Note that while this bears some correspondence to the physical situation of poly(AT) chains with the two hydrogen bonds per base pair, the model will have to deal (as in the original

PBD case) with both AT and GC pairs parametrically, i.e., by adjusting the parameters of the Morse potential to stronger GC bonding while taking account of the combined effects of hydrogen bonds, stacking, and solvent.

The total configurational energy of the proposed system will be composed of two functionally identical, independent—and therefore separable—parts

$$H_{tot}^P = H^P(\{\eta_i\}) + H^P(\{\eta'_i\}),$$

where η_i, η'_i , denote displacements of two independent one-dimensional oscillators along mutually orthogonal directions in the base plane,

$$H^P(\{\eta_i\}) = \sum_{i=1}^{N-1} W(\eta_i, \eta_{i+1}) + \sum_{i=1}^N V(\eta_i)$$

is the potential energy part of the PBD Hamiltonian, and

$$W(\eta_1, \eta_2) = \frac{1}{2}k[1 + \rho e^{-b(\eta_1 + \eta_2)}](\eta_1 - \eta_2)^2$$

represents the nonlinear stacking interaction, parametrized by the stacking range $1/b$, the residual stacking strength k of the high-temperature state, and the ratio $\rho \gg 1$ of the stacking strengths of double vs single-stranded DNA. It should be noted that, as in the original PBD formulation, the on-site potential $V(\eta)$ lumps together effects of hydrogen bonding, stacking, and solvent. Furthermore, the fact that the particles of the second chain have a higher mass changes the kinetic energy but does not enter the configurational energy. The chains are, insofar as configurational partition functions are concerned, identical.

The configurational partition function Z_N^{tot} of the total system composed of N base pairs will thus be equal to the square of the original partition function Z_N as used in, e.g., [1,5]. Explicitly, for $N \gg 1$ and homogeneous systems

$$Z_N^{\text{tot}} = Z_N^2,$$

where

$$Z_N \sim \Lambda_0^{N-1} I_0^2$$

$$I_0 = \int_{-\infty}^{\infty} d\eta e^{-V(\eta)/(2k_B T)} \psi_0(\eta), \quad (5)$$

and Λ_0, ψ_0 are, respectively, the largest eigenvalue and the corresponding eigenfunction of the transfer integral (TI) operator, T is the absolute temperature, and k_B the Boltzmann constant. Details of notation of TI thermodynamics and generalization to inhomogeneous systems are given in [5,9,14].

Hence, all extensive thermodynamic quantities, e.g., enthalpies and entropies and, in particular, the melting entropy as calculated in [5], should be multiplied by a factor of two. Intensive quantities, e.g., the melting fraction, remain the same.

C. Equivalence with a planar vector model

The geometry of the base plane (cf Fig. 1) suggests that we may treat the pair (η_i, η'_i) as a two-dimensional vector. The model may then be thought of as a single discrete planar vector chain with anisotropic local nonlinearity (the decoupled Morse potentials in two independent directions on the

base plane) and an anisotropic nonlocal nonlinear interaction (the nonlinear part of the base stacking term acting in both independent directions). In this version of the theory, because of the particular form of the anisotropy, the dynamics in the two directions become separable and the partition function factorizes.

A planar vector model with strong anisotropy, both local and nonlocal, appears *a posteriori* better adapted to the symmetry of DNA, independently of the derivation presented here in terms of a particular “toy model.” This alternative formulation might also be useful as a starting point for other, more realistic, less anisotropic vector models of DNA melting. Regarding the local anisotropy, the approach is in fact not new; one-dimensional Morse potentials have been successfully used, e.g., to describe melting of hydrogen bonds in the context of self-consistent phonon theory [15]. Deviations from the extreme anisotropic case of the present model are more likely to originate in the nonlocal coupling; however, judging from the results presented in the next sections, I expect cross-correlation effects arising from them to be weak.

In this work, I will refer, equivalently, to either a “two-chain” or a “vector” generalization of the PBD model.

III. THE PBD ENTROPY OF MELTING

A. Numerical estimate in the general, heterogeneous case

The method for calculating the thermodynamic functions is described in [5]. The model parameters, as used in that work, are $b = 0.2 \text{ \AA}^{-1}$, $k = 0.45 \times 10^{-3} \text{ eV/\AA}^2$, $\rho = 50$, $\alpha_{GC} = 6.9 \text{ \AA}^{-1}$, $\alpha_{AT} = 4.2 \text{ \AA}^{-1}$, and a cutoff $\eta_c = 2 \text{ \AA}$, beyond which a PBD base pair is considered open; furthermore, the values $D_{GC} = 0.1655 \text{ eV}$, $D_{AT} = 0.1255 \text{ eV}$, which were used to fit the melting profile of PBR322 (Eco RI cut) [7], serve as reference values at salt concentration 0.075 M.

The melting fraction and the entropy per site, calculated using the above set of parameters and the full genomic sequence of PBR322 [5], are shown in the onset of Fig. 2; appropriately rescaled, the two curves are essentially identical. In order to obtain a numerical estimate of the melting entropy, I make use of the Ansatz

$$\frac{ds_1(T)}{dT} = \Delta s^* \frac{d\theta}{dT}, \quad (6)$$

where $s_1(T)$ is the entropy per site at temperature T in units of k_B , and $d\theta/dT$ the temperature derivative of the melting fraction, both computed for a scalar chain; it is possible to obtain a theoretical estimate of the melting entropy per site $\Delta s^* = 4.40$. The procedure is shown in Fig. 2. Since the numerics is done for a scalar PBD chain, the end result for the vector model is

$$2 \times \Delta s^* = 8.8. \quad (7)$$

B. Elastic entropy

Neither the PBD model in its original Hamiltonian form nor its vector generalization can describe entropic elasticity. Each base pair may dissociate, in the sense that the base pair separation coordinates exceed a certain value, yet, strictly, continues to reside on the base plane and is not free to sample three-dimensional space. The PBD “molten” state can

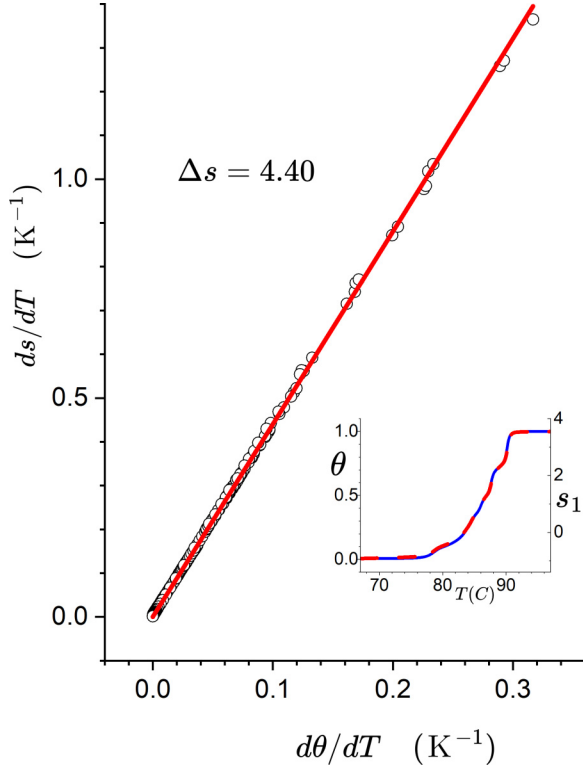


FIG. 2. Melting entropy for a single PBD chain. The onset shows that the curves for the entropy s_1 (right y axis, red, dashed) and the melting fraction (left y axis, solid, blue) θ coincide when appropriately rescaled. The main plot shows the temperature derivative of the anharmonic part of the entropy vs the temperature derivative of the melting fraction (open circles). The linear fit to the data through the origin yields the slope Δs^* . The absence of scattering in the data points (which originate in different regions of the melting profile) confirms the validity of the Ansatz (6) and, hence, that the melting entropy of a base pair is largely independent of chemical composition and/or position in the sequence [7].

be thought of as a longitudinally “straight” ladder with very wide steps and no overall bending. The change in entropy which originates in the different embedding of such a stiff polymer with very small orientational entropy vs two soft single-stranded polymer chains in 3D-space can in fact be estimated separately, as follows.

A straightforward way to compute the elastic entropy of a polymer composed of N monomers, monomer distance a , and persistence length λ is to approximate the polymer by a freely jointed chain (FJC) consisting of $N/(2\lambda/a)$ stiff segments, each of which has a length 2λ (Kuhn length). Each segment contributes a factor 4π to the conformational partition function

$$\Omega = (4\pi)^{aN/(2\lambda)},$$

leading to an elastic entropy of the polymer, in units of k_B ,

$$S_{el} = \ln \Omega = \ln(4\pi)^{aN/(2\lambda)}.$$

The orientational (elastic) melting entropy of DNA per base pair can then be estimated as the difference

$$\Delta s_{el} \equiv \frac{\Delta S_{el}}{N} = \ln(4\pi) \left\{ 2 \frac{a'}{2\lambda'} - \frac{a}{2\lambda} \right\}, \quad (8)$$

where a, a', λ, λ' are, respectively, the monomer and persistence lengths of double- and single stranded chains, and the factor of two reflects the two separate strands of the denatured state. Using standard values $a = 0.34$ nm, $a' = 0.63$ nm [16], $\lambda = 53.4$ nm [17], and $\lambda' = 1.62$ nm [18], results in a value $\Delta s_{el} = 0.98$ per base pair.

It should be noted here that the second term in (8) is very small, as expected; the elastic entropy of double-stranded DNA is essentially zero. The main contribution to Δs_{el} comes from the first term and is sensitive to the choice of parameters. I have chosen the carefully averaged SAXS value of the monomer distance reported in [16]. The value of λ' used was interpolated from results of [18] for a salt concentration of 0.075 M, consistent with the data used in the previous section; lower values were reported at higher salt concentrations and vice versa.

Combining the above estimate of 3D elastic entropy with the result of the previous section, I obtain a value of

$$\Delta s = 2 \times \Delta s^* + \Delta s_{el} = 9.78. \quad (9)$$

This is roughly in line with experiments reporting a melting entropy averaged over many calorimetric experiments on long polymers, $\Delta s_{exp} = 12.5 \pm 0.9$ [7], and the somewhat lower estimate, $\Delta s_{exp} = 11.0$ of [8], based on a systematic analysis of oligomer data.

IV. UNZIPPING

Calculating the transverse force required to separate the two DNA strands is quite straightforward in the homogeneous limit within the PBD model [9]. In the presence of such a force, the Hamiltonian is modified by a term $-f\eta_N$, where η_N is the displacement coordinate of the last base pair. This modifies one of the two end integrals I_0 in Eq. (5), introducing an extra factor $e^{f\eta/(k_B T)}$ in the integrand. Now it is known from the continuum limit of the PBD model [3] that the eigenfunction ψ_0 decays exponentially at large values of its argument. This property survives in the discrete case as well, as shown in Fig. 3, i.e., $\psi_0(\eta) \propto e^{-\eta/\xi_\perp}$, where $\xi_\perp$ is a characteristic length scale. As the force increases, an instability (divergence of I_0) occurs at a value

$$f^* = \frac{k_B T}{\xi_\perp}.$$

Within the vector model introduced here, the unzipping force is equal to $2f^*$, since half of the physical external force must be applied to each chain. The result, for the parameters of Sec. III (corresponding to a salt concentration of 75 mM) is an unzipping force of 10 pN for AT and 17.6 pN for GC at 300 K, i.e., an average of 13.8 pN. In comparison, Ref. [10] reports 13 pN at 60% GC content and 150 mM salt concentration, whereas separately measured unzipping forces of 9 ± 3 pN and 20 ± 3 pN have been reported for AT and GC, respectively [19], also at 150 mM salt concentration.

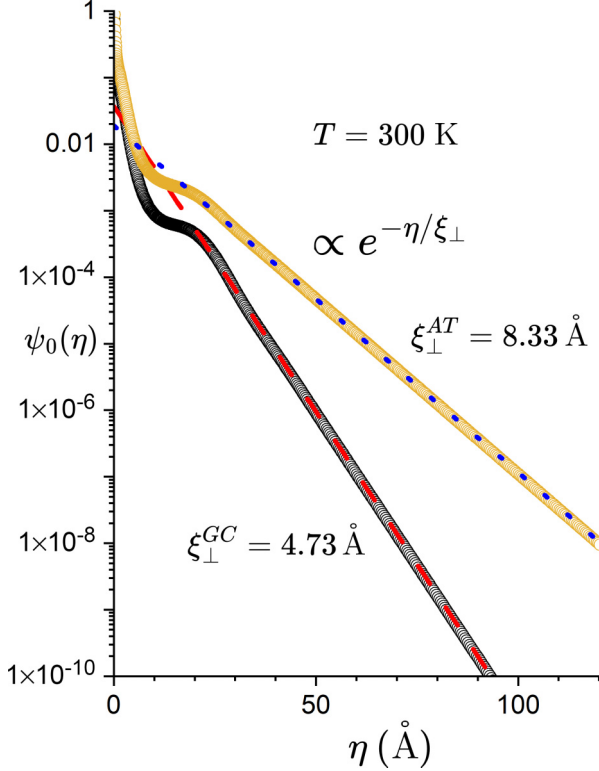


FIG. 3. The eigenfunction ψ_0 in the homogeneous case. Note the purely exponential asymptotics which allow the extraction of $\xi_\perp$ for GC and AT. Parameters are the same as in Sec. III.

V. HAIRPIN OPENING RATES

Single strands of DNA with complementary ends may close on themselves to form hairpinlike structures. The complementary ends form the stem, which is effectively a double-helical oligomer, while the rest of the chain arranges itself in a looplike fashion. Raising the temperature results in dissociation of the hairpin structure via melting of the stem. The melting curves in the limiting case of a short stem and an intermediate-size T loop [20] can be well understood [9] in terms of the stem nearest-neighbor enthalpies and entropies given in [21] and a self-avoiding walk polymer model of the loop. The complete thermodynamic and kinetic problem is a complex one, involving both DNA and polymer statistical mechanics, diffusion properties of ss-DNA, as well as intermediate states formed during trapping by mismatches in the process of folding [22,23]. In this section, I will consider measurable quantities for which such complexities are either absent or they cancel out, e.g., in ratios of reaction rates. In such simple cases, hairpin opening occurs by melting [20,24] or “unzipping” [25,26] of the stem; the barrier for the opening transition is determined by the stem’s free energy difference relative to the open state [25], which contains both entropic and enthalpic contributions, i.e., the opening rate is

$$k_{op} = \gamma_0 e^{\Delta S_{stem}/k_B} e^{-\Delta H_{stem}/(k_B T)}, \quad (10)$$

where γ_0 is controlled by the diffusion rate of ss-DNA and can be determined by the measured closing rate [24] to be of the order of $15 \mu\text{sec}^{-1}$.

In the case of the 31-base sequence $5' - \text{CCCAA} - (\text{T})_{21} - \text{TTGGG} - 3'$, the transition enthalpy has been determined to be equal to 32 Kcal/mol [24]. This value can be understood in terms of the nearest-neighbor DNA melting enthalpies [21],

$$\begin{aligned} \Delta H_{stem} &= \Delta H_{CoG}^{init} + 2\Delta H_{CC/GG} + \Delta H_{CA/GT} \\ &\quad + \Delta H_{AA/TT} + \Delta H_{AcT}^{init} \\ &= -0.1 + 2 \times 8 + 8.5 + 7.9 - 2.3 = 30 \text{ Kcal/mol}, \end{aligned}$$

where initiation values have been used for the stem end points.

Alternatively, in the framework of the PBD model, the transition enthalpy can be estimated by assigning enthalpies to single base pairs. The values which enter such estimates can be obtained by computing the thermodynamics of long ($N=1000$) sequences of pure AT or pure GC. I use the parameters introduced in Sec. III, modified for salt concentration 0.2 M according to [5]. This results in a slight shift of the Morse potential depths, $D_{GC} = 0.1291$ eV and $D_{GC} = 0.1681$ eV, respectively. The relevant results of TI thermodynamics are displayed in Fig. 4, from which one extracts transition enthalpies $\epsilon_{GC}^{(0)} = 0.1545$ eV, $\epsilon_{AT}^{(0)} = 0.1190$ eV and transition entropies $s_{GC}^{(0)} = 4.6454$, $s_{AT}^{(0)} = 4.0759$, respectively. Including the elastic entropy contribution, this results in total entropies per base pair $s_{GC} = 10.27$, $s_{AT} = 9.13$. The corresponding melting temperatures are $T_{GC}^m = \epsilon_{GC}^{(0)}/s_{GC}^{(0)} = 386.10$ K and $T_{AT}^m = \epsilon_{AT}^{(0)}/s_{AT}^{(0)} = 338.82$ K and agree with those determined directly from the profiles of Fig. 4 within 0.03 K.

In order to maintain consistency of the elastic-entropy-corrected thermodynamics with the melting profiles, it is necessary to correct the single chain GC and TA enthalpies by amounts $T_{GC}^m \Delta s_{el}/2$ and $T_{AT}^m \Delta s_{el}/2$, respectively. The resulting total transition enthalpies per base pair for the vector model are $\epsilon_{GC} = 0.3417$ eV, $\epsilon_{AT} = 0.2666$ eV, respectively. I note in passing that these values are within 3% of the respective Morse potential depths multiplied by a factor of two.

The total enthalpy of the stem can now be estimated as

$$\Delta H_{stem} = 3\epsilon_{GC} + 2\epsilon_{AT} = 1.5583 \text{ eV} = 35.9 \text{ Kcal/mol},$$

or 60.7 in units of $k_B T$ at 25 C, about 12% higher than the experimentally determined value. This is consistent with the difference between the enthalpies derived in this work and the somewhat lower effective enthalpy parameters 0.2910 and 0.2465, for GC and AT, respectively [11], which fit the experimental data of [24].

As a second hairpin example, where now entropic stem effects are important, I will consider the case of two otherwise identical hairpins, where the ten-basepair long stems differ only at a single site, i.e., GAAGAGGGAG vs GAA-GAGGGGG. In this case the ratio of the opening rates has been measured at 25 C to be 5.6 ± 1.5 [25]. The substitution of a single GC by an AT pair produces, in general, both an energy and an entropy difference. Using the nearest neighbor model with the standard parameter set [21] results in zero enthalpic change and a $\Delta S^* = 3.4$ cal/K/mol. The opening rate ratio is estimated from (10) as $e^{\Delta S^*/k_B} = 5.53$, in excellent agreement with experiment.

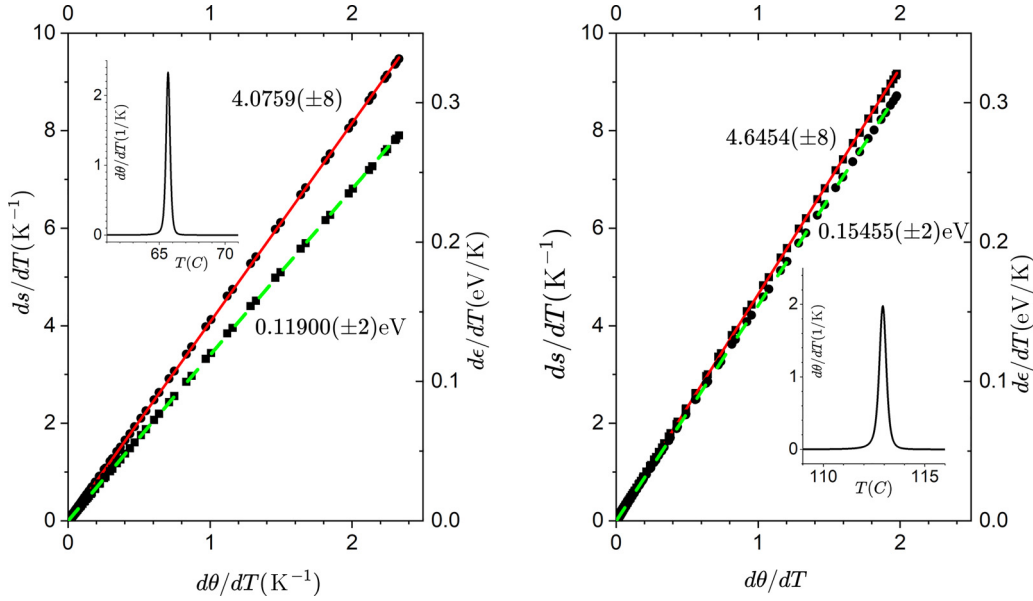


FIG. 4. Extraction of single chain PBD ($N = 1000$) average energies and entropies per site (cf. Fig. 2). Left panel: pure AT ds/dT as a function of the temperature derivative of the melting fraction (squares, left y axis); $d\epsilon/dT$ as a function of the temperature derivative of the melting fraction (circles, right y axis); the lines represent least square fits and their slopes provide estimates for melting entropy and energy, respectively; the onset shows the temperature derivative of the melting fraction as a function of temperature. Right panel: same for pure GC. The Morse depth parameters used, $D_{AT} = 0.1291$ eV, $D_{GC} = 0.1681$, correspond to a salt concentration 0.2 M.

In the vector PBD case, the rate ratio will be

$$e^{(s_{AT}-s_{GC})} e^{-(\epsilon_{AT}-\epsilon_{GC})/(k_B T)} = e^{-1.14+2.92} = 5.93, \quad (11)$$

also in line with experiment.

In order to address the discrepancy of PBD opening rates by several orders of magnitude compared to experiment, a problem raised in Ref. [11], one must return to (10) and consider the entropic term as well. In the case of the five-stem hairpin considered above, the stem's entropic contribution within the vector PBD framework is

$$\Delta S_{\text{stem}}/k_B = 3s_{GC} + 2s_{AT} = 49.1.$$

At a temperature of 25 C, this results in a free energy difference $60.7 - 49.1 = 11.6$ in units of $k_B T$, and an opening rate $k_{op}/\gamma_0 = e^{-11.6} = 9.17 \times 10^{-6}$. Had I used the single-chain PBD model, the result would have been 5.2×10^{-3} , i.e., two and a half orders of magnitude higher; this is exactly the order of discrepancy between PBD and experimental values exhibited in Fig. 6 of Ref. [11].

As the stem length increases to $N_{\text{stem}} = 30$ base pairs, the discrepancy between single-chain PBD opening rates predictions and experimental values rises up to ca. 17 orders of magnitude (cf. Fig. 4 of Ref. [11]). At 50% GC content and 25 C temperature, the average values of vector PBD melting energy and entropy per site are $\epsilon = 0.3042$ eV $= 11.95 k_B T$ and $s = 9.70$, respectively. The opening rate is then expected to be proportional to $b^{N_{\text{stem}}}$ where $b = \exp\{-\epsilon/(k_B T) + s\} = \exp(-11.95 + 9.70) = \exp(-2.25) = 0.105$. The corresponding number extracted from experiment [26] in the analysis of [11] is $b = 0.095$, i.e., lower by approximately 10%. For a stem length equal to 30, this will result in a discrepancy of $(0.105/0.095)^{30} = 20$, i.e., the opening rate calculated within the vector PBD model will still be higher than the observed one by a factor of 20. While not negligible,

this is certainly a long way from the 17 orders of magnitude discrepancy generated by the single-chain PBD model.

VI. DISCUSSION

This paper has argued that a reductionist model of DNA melting according to the logic of Ref. [1] can follow naturally from a microscopic picture of the base-pair plane of the double helix in which carbonyl and exocyclic as well as endocyclic amino groups appear explicitly. The interest distinctive feature of the particular reductionist scheme is that it leads to two independent, nonlinear oscillators per base pair, which, in the presence of the stacking interaction, generate two independent chains of coupled nonlinear oscillators, thus doubling the magnitudes of extensive thermodynamic quantities while leaving melting profiles as computed in, e.g., [5], unchanged. The analysis of diverse types of data, e.g., melting entropy, unzipping forces, and hairpin opening rates, presents rather compelling evidence in favor of the vector PBD model introduced in this work. However, it should be noted that this type of analysis simply shows a very complete thermodynamic equivalence between a Hamiltonian model with a particular set of seven parameters and the nearest-neighbor model with its set of ten enthalpies and ten entropies. There is as yet little specific experimental evidence in favor of the microscopic physics underlying the PBD model, either in the vector or in its original scalar version. On a somewhat speculative vein, one might therefore consider the following problem.

NMR experiments with DNA dodecamers using imino proton exchange suggest that DNA base pairs open spontaneously at room temperatures, one at a time [13]. The equilibrium constant between open and closed states of a base pair has been measured to lie between $10^{-5} - 10^{-6}$, the lower val-

ues corresponding to GC and the higher to AT pairs. For comparison, a PBD-based long chain calculation performed with the parameters of Sec. III at room temperatures yields melting fraction values $\theta = 0.7 \times 10^{-3}$ (GC) and 2.5×10^{-3} (AT), which are two to three orders of magnitude larger. Setting aside for a moment the question of whether long chain results can be applied to small systems (the example of hairpins shows they—approximately—can), one may pause to examine the measurement process. The typical experimental situation in calorimetry or UV absorption, is that the disruption of a single hydrogen bond (as opposed to a full opening of a base pair) at any number of sites is detected as proportionately contributing to the melting fraction. Similarly, the melting fraction computed in the PBD model is the fraction of sites where the spatial separation coordinate exceeds a certain value; adding an uncorrelated identical chain does not change the computed value of θ . In contrast, imino proton exchange requires a fully open state of the base pair, i.e., in the context of the present paper, a simultaneously open i th site on both chains (or, in the vector language, both independent components of the vector order parameter exceeding a certain value). The probability for such an event to occur is θ^2 , which at room temperatures has a much lower value, namely, 0.5×10^{-6} and 6×10^{-6} for pure GC and AT, respectively. These values are close to the equilibrium constants determined in [13].

In this context, it is interesting to note that, in order to obtain such low values of the open state probability with the same parameters in a standard PBD calculation, a much

higher value of the cutoff $\eta_c = 5$ Å would be needed [5]; that result already indicated that the open state of imino proton exchange was probably not identical with what calorimetry or UV absorption detects. The model presented here clarifies this difference in a more satisfactory manner.

The removal of the serious discrepancy between theory and experiment on the opening rate of hairpins has an important, perhaps not immediately obvious, consequence regarding further research. Usage of the PBD model in mesoscopic molecular dynamics simulations of dynamical properties, as has been done in [11], has been explicitly validated; as a result, the interpretation of PBD simulation data on dynamical properties has become less speculative. It would be interesting to explore this, e.g., in order to provide an up-to-date theoretical estimate of the average lifetime of a base pair [13].

ACKNOWLEDGMENTS

I acknowledge thoughtful comments and critical questions raised by both referees.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

- [1] M. Peyrard and A. R. Bishop, Statistical mechanics of a non-linear model for DNA denaturation, *Phys. Rev. Lett.* **62**, 2755 (1989); T. Dauxois, M. Peyrard, and A. R. Bishop, Entropy driven DNA denaturation, *Phys. Rev. E* **47**, R44 (1993).
- [2] N. Theodorakopoulos, T. Dauxois, and M. Peyrard, Order of the phase transition in models of DNA thermal denaturation, *Phys. Rev. Lett.* **85**, 6 (2000).
- [3] T. Dauxois, N. Theodorakopoulos, and M. Peyrard, Thermodynamic instabilities in one dimension: Correlations, scaling and solitons, *J. Stat. Phys.* **107**, 869 (2002).
- [4] D. Cule and T. Hwa, Denaturation of heterogeneous DNA, *Phys. Rev. Lett.* **79**, 2375 (1997).
- [5] N. Theodorakopoulos, Melting of genomic DNA: Predictive modeling by nonlinear lattice dynamics, *Phys. Rev. E* **82**, 021905 (2010).
- [6] A. Wildes, N. Theodorakopoulos, J. Valle-Orero, S. Cuesta-López, J.-L. Garden, and M. Peyrard, The thermal denaturation of DNA studied with neutron scattering, *Phys. Rev. Lett.* **106**, 048101 (2011); Structural correlations and melting of B-DNA, *Phys. Rev. E* **83**, 061923 (2011).
- [7] S. G. Delcourt and R. D. Blake, Stacking energies in DNA, *J. Biol. Chem.* **266**, 15160 (1991).
- [8] J. SantaLucia, H. T. Allawi, and P. A. Seneviratne, Improved nearest-neighbor parameters for predicting DNA duplex stability, *Biochemistry* **35**, 3555 (1996); This work, mainly on oligopolymer structures, explicitly validates the result of [7] on the value of the average melting entropy of long polymers.
- [9] N. Theodorakopoulos, *The Statistical Physics of DNA* (World Scientific, Singapore, 2020).
- [10] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, Mechanical separation of the complementary strands of DNA, *Proc. Natl. Acad. Sci. USA* **94**, 11935 (1997).
- [11] O. Dahlen and T. S. van Erp, Mesoscopic modeling of DNA denaturation rates: Sequence dependence and experimental comparison, *J. Chem. Phys.* **142**, 235101 (2015).
- [12] M. Marty-Roda, O. Dahlen, T. S. van Erp, and S. Cuesta-López, Improving the mesoscopic modeling of DNA denaturation dynamics, *Phys. Biol.* **15**, 066001 (2018).
- [13] M. Gueron, M. Kochoyan, and J.-L. Leroy, A single mode of DNA base-pair opening drives imino proton exchange, *Nature (London)* **328**, 89 (1987); J.-L. Leroy, M. Kochoyan, T. Huynh-Dinh, and M. Gueron, Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton, *J. Mol. Biol.* **200**, 223 (1988).
- [14] N. Theodorakopoulos, Bubbles, clusters and denaturation in genomic DNA: Modeling, parametrization, efficient computation, *J. Nonl. Math. Phys.* **18**, 429 (2011).
- [15] Y. Gao and E. W. Prohofsky, A self-consistent phonon theory of hydrogen bond melting, *J. Chem. Phys.* **80**, 2242 (1984).
- [16] M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, Probing single-stranded DNA conformational flexi-

- bility using fluorescence spectroscopy, *Biophys. J.* **86**, 2530 (2004).
- [17] C. Bustamante, J. F. Marko, E. D. Siggia, and S. B. Smith, Entropic elasticity of Lambda phage DNA, *Science* **265**, 1599 (1994).
- [18] A. Bosco, J. Camunas-Soler, and F. Ritort, Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions, *Nucleic Acids Res.* **42**, 2064 (2014).
- [19] M. Rief, H. Clausen-Schaumann, and H. E. Gaub, Sequence-dependent mechanics of single DNA molecules, *Nat. Struct. Biol.* **6**, 346 (1999).
- [20] N. L. Goddard, G. Bonnet, O. Krichevsky, and A. Libchaber, Sequence dependent rigidity of single stranded DNA, *Phys. Rev. Lett.* **85**, 2400 (2000).
- [21] H. T. Allawi and J. SantaLucia, Thermodynamics and NMR of internal GT mismatches in DNA, *Biochemistry* **36**, 10581 (1997).
- [22] J. Errami, M. Peyrard, and N. Theodorakopoulos, Modeling DNA beacons at the mesoscopic scale, *Eur. Phys. J. E* **23**, 397 (2007).
- [23] Y. Shen, S. V. Kuznetsov, and A. Ansari, Loop dependence of the dynamics of DNA hairpins, *J. Phys. Chem. B* **105**, 12202 (2001).
- [24] G. Bonnet, O. Krichevsky, and A. Libchaber, Kinetics of conformational fluctuations in DNA hairpin-loops, *Proc. Natl. Acad. Sci. USA* **95**, 8602 (1998).
- [25] 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).
- [26] 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. USA* **103**, 6190 (2006).