Experimental tests of the Peyrard-Bishop model applied to the melting of very short DNA chains

Alessandro Campa[1,*] and Andrea Giansanti[2]
[1]*Physics Laboratory, Istituto Superiore di Sanità and INFN Sanità, Viale Regina Elena 299, 00161 Roma, Italy*
[2]*Physics Department, Università di Roma ''La Sapienza'' and INFM Unità Roma 1, Piazzale Aldo Moro 2, 00185 Roma, Italy*
(Received 23 February 1998)

The melting curves of short heterogeneous DNA chains in solution are calculated on the basis of statistical thermodynamics, and compared to experiments. The computation of the partition function is based on the Peyrard-Bishop Hamiltonian, which has already been adopted in a theoretical description of the melting of long DNA chains. In the case of short chains it is necessary to consider not only the breaking of the hydrogen bonds between single base pairs, but also the complete dissociation of the two strands forming the double helix. [S1063-651X(98)02709-3]

PACS number(s): 87.10.+e, 64.70.−p, 05.90.+m

I. INTRODUCTION

There is a need for a theory of the melting of short DNA chains (oligonucleotides). The melting is the highly cooperative thermal disruption of hydrogen bonds between complementary bases in the double helix, as usually monitored by the UV absorption increment due to the unstacking of the separated bases [1]. At the equilibrium melting temperature, half of the bonds are disrupted. Synthetic oligonucleotides of a fixed length and base pairs sequence have been used for a long time as model systems for the study of the structural and thermodynamical properties of the longer and more complex natural forms of DNA [2]. Many studies have shown the effects of both sequence and solvent composition on the melting curves of oligonucleotides in solution [3]. More recently, particular attention has been given to the study of the sequence effects on the thermal stability of a variety of specially designed oligonucleotides, due to their importance in the exploitation of molecular biological techniques in gene therapy [4] and genome mapping [5]. Predictive information has been gained through an extensive thermodynamical investigation of the melting behavior of oligonucleotides, based on the computation of the Gibbs free energy, at a fixed solvent composition, as a sum of contributions from nearest neighbors in the sequences [6,7]. This phenomenology, and the predictive power of the thermodynamical approach, should then be confronted with a microscopic theory of short, heterogeneous DNA chains.

Modeling of DNA melting was initially motivated by the study of the important process of transcription, in which the double helix has to be opened locally to allow a reading of the genetic code. This was based, many years ago, on Ising-like models [8,9], and more recently on an approach based on the modified self-consistent phonon approximation [10] (see also Ref. [11] and references therein). These methods allow only equilibrium estimates of the probability of bond disruption. However, it is also important to consider DNA dynamics, both at melting and premelting temperatures. There is an interest in relaxation and kinetic phenomena, which are relevant for pharmacological applications [4], and in the study of nonlinear energy localization and transduction. With a particular focus on the last problem, discrete nonlinear models of DNA (see, e.g., Refs. [12,13], and, for a review, Ref. [14]) have been introduced; sequence effects were considered in Ref. [15]. These models are appealing, because they are simplified microscopic models with a small number of degrees of freedom, and thus are also affordable for the simulation of very long times. The experimentally available melting curves offer a way to optimize the parameters of these models, and therefore also increase confidence in their use in dynamical studies.

With a particular interest in thermal stability, a dynamical model was introduced by Peyrard and Bishop in 1989 [16] (the PB model). The authors showed, through statistical mechanics calculations and constant temperature molecular dynamics [16–18], applied to the case of a very long homogeneous DNA chain, that the model can give a satisfactory melting curve, especially after the improvement introduced in Ref. [18]. The PB model has been successively applied to heterogeneous chains, either modeling the heterogeneity with a quenched disorder [19], or properly choosing basis sets of orthonormal functions for the kernels appearing in the expression of the partition function [20], but comparison with experimental data was not attempted. In all these works the fact that the DNA's considered are quite long was essential, for the following reason. In a solution with two types of DNA single strands A and B, there is a thermal equilibrium between dissociated strands and associated double strands (the duplexes AB), and a thermal equilibrium, in the duplexes, between broken and unbroken interbase hydrogen bonds. The average fraction θ of bonded base pairs can then be factorized as $\theta = \theta_{\rm ext}\theta_{\rm int}$ [8,9]. $\theta_{\rm ext}$ is the average fraction of strands forming duplexes, while $\theta_{\rm int}$ is the average fraction of unbroken bonds in the duplexes. The dissociation equilibrium can be neglected in the case of long chains, where $\theta_{\rm int}$ and thus θ go to 0 when $\theta_{\rm ext}$ is still practically 1. Conversely, in the case of short chains the processes of single bond disruption and strand dissociation tend to happen within the same temperature range; therefore, the computation of both $\theta_{\rm int}$ and $\theta_{\rm ext}$ is essential, as we will point out while presenting our comparison with experimental data. In Ref. [20] the fac-

*Author to whom correspondence should be addressed. Electronic address: campa@axiss.iss.infn.it

torization of θ is stated, but only the case of long chains is then considered.

The aim of this work is to show, through a comparison with experimental data, that the one-dimensional PB model can be used to compute the melting curves of short DNA's. It will also be shown how to take into account the dissociation equilibrium.

II. MODEL AND RESULTS

The potential of the PB model [16–18] is given by

$$U=\sum_i \left\{ \frac{k}{2}[1+\rho e^{-\alpha(y_{i+1}+y_i)}](y_{i+1}-y_i)^2 + D_i(e^{-a_i y_i}-1)^2 \right\}, \quad (1)$$

where y_i is the distance between the ith complementary bases minus their equilibrium separation. The parameters k, ρ, and α refer to the anharmonic stacking interaction, while the interbase bond is represented by a Morse potential, with depth D_i and width a_i. In Refs. [16–18] there is only a single parameter D because only homogeneous DNA's have been considered. The stacking interaction, that in the first attempts [16,17] was purely harmonic ($\rho=0$), decreases when the complementary bases reach farther (ρ positive): this ρ dependent nonlinear term was found to be relevant to give cooperativity to the melting process [18].

To model heterogeneous DNA's, we have inserted two different values of D_i, according to the two possible Watson-Crick base pairs: adenine-thymine (AT) and guanine-cytosine (GC). The former has two hydrogen bonds, while the latter has three. We have then chosen a depth for the GC Morse potential 1.5 times that for the AT Morse potential. The complete set of parameter values that we have chosen is $k=0.025$ eV/Å^2, $\rho=2$, $\alpha=0.35$ Å^{-1}, $D_{\mathrm{AT}}=0.05$ eV, $D_{\mathrm{GC}}=0.075$ eV, $a_{\mathrm{AT}}=4.2$ Å^{-1}, and $a_{\mathrm{GC}}=6.9$ Å^{-1}. For a given set of values, the melting temperature of homogeneous DNA can be deduced with the technique of the transfer matrix method [16–18], and the values have been adjusted to reproduce the experimentally observed melting temperature of long homogeneous DNA in the most usual solvent conditions [9,21]. The set of values given above has been used for all the different oligonucleotides considered. The parameter values are close to those in the original PB model.

We have then made a statistical mechanics computation, in which partition functions have been used to obtain both θ_{int} and θ_{ext}. For the computation of θ_{int} one has to separate the configurations describing a double strand on the one hand, and dissociated single strands on the other. The very possibility of dissociation makes this a nontrivial problem. We have adopted the following strategy. The ith bond is considered disrupted if the value of y_i is larger than a chosen threshold y_0. We therefore have defined a configuration to belong to the double strand if at least one of the y_i's is smaller than y_0. It is then natural to define θ_{int} for an N base pair duplex by

$$\theta_{\mathrm{int}}=\frac{1}{N}\sum_{i=1}^{N}\langle \vartheta(y_0-y_i)\rangle,$$

where $\vartheta(y)$ is the Heaviside step function and the canonical average $\langle\ \rangle$ is defined considering only the double strand configurations. We have chosen a value of 2 Å for y_0. After a discretization of the coordinate variables and the introduction of a proper cutoff on the maximum value of the y_i's [10], the computations needed for the canonical averages are readily reduced to the multiplication of finite matrices, since the potential (1) couples only nearest neighbors, and are easily performed by suitable computer programs.

Let us now consider θ_{ext}. At equilibrium the chemical potentials of the three species A, B, and AB [22] are related by the equation: $\mu_{AB}-\mu_A-\mu_B=0$. Using the definition of the chemical potentials as derivatives of the free energy, and in turn the relation of the latter to the partition functions, we obtain an equation involving appropriate partition functions. In the usual experimental conditions the solutions can be considered ideal; with the further assumption that the model effectively takes the presence of the solvent into account, we obtain the usual equilibrium condition:

$$\frac{N_{AB}Z(A)Z(B)}{N_A N_B Z(AB)}=1,$$

where N_j is the number of molecules of species j in the volume V considered, and $Z(j)$ is the partition function of a molecule of species j in V [23]. The numbers N_j are related by the constraints $2N_{AB}+N_A+N_B=\mathrm{const}\equiv 2N_0$ and $\Delta N_A=\Delta N_B=-\Delta N_{AB}$. Considering the case $N_A=N_B$ (the experimental curves that we are presenting are made in these conditions, with the duplex obtained by annealing equal concentrations of A and B), we arrive at the following expression for $\theta_{\mathrm{ext}}\equiv N_{AB}/N_0$:

$$\theta_{\mathrm{ext}}=1+\delta-\sqrt{\delta^2+2\delta},$$

where δ is given by the expression:

$$\delta=\frac{Z(A)Z(B)}{2N_0 Z(AB)}\equiv\frac{Z_{\mathrm{int}}(A)Z_{\mathrm{int}}(B)}{a_{\mathrm{av}}Z_{\mathrm{int}}(AB)}\,\frac{a_{\mathrm{av}}Z_{\mathrm{ext}}(A)Z_{\mathrm{ext}}(B)}{2N_0 Z_{\mathrm{ext}}(AB)}, \quad (2)$$

where in the rightmost side we have introduced the separation of the partition functions in internal and external parts [8,9]; the meaning of a_{av} will be explained in a moment. For the calculation of the internal functions, that do not include the overall translation of the molecules, we use the DNA model described above (which is also simply adapted to the description of single strands, allowing an analytical evaluation: only a harmonic stacking interaction remains, which is weaker than in the duplex, since in this case the term involving ρ is 0). We have chosen to insert $a_{\mathrm{av}}=\sqrt{a_{\mathrm{AT}}a_{\mathrm{GC}}}$ in the last side of Eq. (2) to make both fractions separately dimensionless; therefore they cannot depend on the choice of units. Without any such normalization the first fraction would have the dimensions of an inverse of a length, since the overall translation is not included in Z_{int}. It is included in the external functions, that, however, also have to take into account the dynamics not described by the simple one-dimensional model, related to conformational movements (like, for ex-

ample, the winding of the strands). This point was already considered in Ising models: the influence on the dissociation process of the degrees of freedom not described by the model cannot be neglected, and it must be accounted for in some way. In analogy to what was proposed for the Ising models [8,9] on the basis of the partition functions of rigid bodies [24], we make the following choice:

$$\frac{a_{\mathrm{av}} Z_{\mathrm{ext}}(A) Z_{\mathrm{ext}}(B)}{2 N_0 Z_{\mathrm{ext}}(AB)} = \frac{n^*}{n_0} N^{-p\theta_{\mathrm{int}}+q}, \tag{3}$$

where the parameters p and q can be fixed by a comparison with experimental melting curves; n_0 is the single strand concentration N_0/V, and n^* is a chosen reference concentration (we have taken 1 μM, a usual concentration in experiments). We defer further comments about this equation until after the presentation of the results.

Here we show the comparison of our calculations with the experimental melting curves that have been obtained, in our lab, for three different oligonucleotides, in a 10-mM Na phosphate buffer, 0.1-mM Na_2EDTA and 200-mM NaCl, *p*H 6.7. One of the oligonucleotides contained 27 base pairs, and the other two had 21 base pairs. The sequences are given by

$$(s_1)\ {}^{5'}\mathrm{CTTCTTATTCTTATTGTTCGTCTTCTC}_{3'},$$

$$(s_2)\ {}^{5'}\mathrm{CTCTTCTCTTCTTTCTCTCTC}_{3'},$$

$$(s_3)\ {}^{5'}\mathrm{GTGTTAACGTGAGTATAGCGT}_{3'},$$

and by the respective complementary strands. We have considered the case s_3 at two different concentrations. The single strand concentration was (s_1) 2.4 μM, (s_2) 1.7 μM, and (s_3) 3.1 and 120 μM. In Fig. 1 we show the experimental and computed melting curves. Our computations, as stated above, take into account both factors θ_{int} and θ_{ext}. To assess quantitatively the effect of θ_{ext}, we have checked that at complete melting, when θ becomes practically zero (i.e., $\phi = 1$), the values of θ_{int} are still 0.81 (a), 0.85 (b), 0.83 (c), and 0.78 (d). This is consistent with the Ising-like computation shown by Wartell and Benight for a 95 base pair DNA, where θ_{int} at complete melting is still 0.63 (see Fig. 8 of Ref. [9]). As can be seen, there are sequence and concentration effects on the experimental melting curves, which are well reproduced by the computed curves. Note that a 40-fold concentration increase for s_3 yields an increase of only 5° in the melting temperature (a logarithmic dependence on the concentration is expected [1]). Similar differences between curves at the low concentrations should then be due to sequence and length effects. We would like to stress that in the case s_3 the parameters p and q have been fitted to the experimental curve at the lower concentration. The comparison with the experimental curve at the higher concentration has then been performed with only the change of the value of n_0 in Eq. (3), without changing the values of p and q; this has reproduced the difference between the melting temperatures of the two cases, that differ by about 5°. This fact indicates that the concentration dependence of the left-hand side of Eq. (3) is described by the preexponential factor, while the parameters p and q are related to the molecular conformation. At the beginning of the melting region, especially Fig. 1(a) shows a slight disagreement between computed and experimental curves. However, it has to be noted that our experimental curves have been normalized following a standard procedure, in which base line subtraction can have minor effects on the shape of the normalized curves, particularly in the premelting region (see, e.g., Sec. 1 of Ref. [9]).

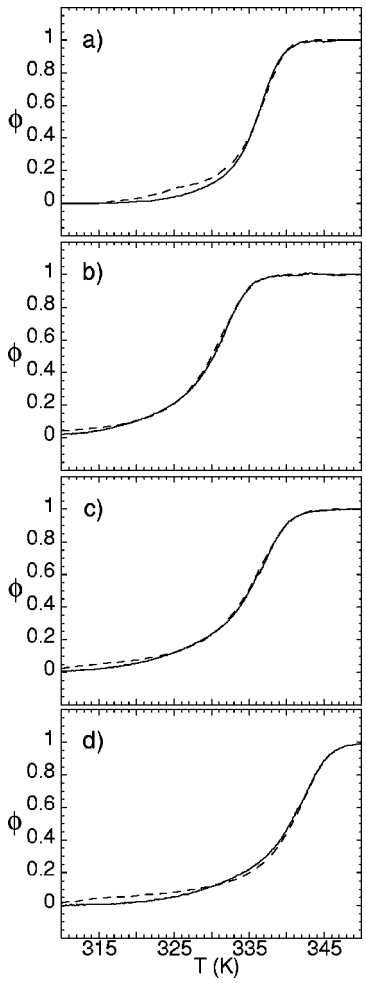

FIG. 1. Experimental melting profiles (full lines) and theoretical results (dashed lines) for the three DNA chains. We have plotted the value of $\phi \equiv 1 - \theta$. (a) Sequence s_1. (b) Sequence s_2. (c) Sequence s_3 at the lower concentration. (d) Sequence s_3 at the higher concentration. The fitted parameters p and q have the following values: $p = 32.43$ and $q = 29.30$ for s_1, $p = 36.77$ and $q = 34.89$ for s_2, and $p = 29.49$ and $q = 27.69$ for s_3.

III. CONCLUSIONS

In conclusion, our comparisons show that it is feasible to compute the equilibrium melting profile of DNA oligonucleotides with the PB nonlinear model. We would also like to note that the modelization of the external partition functions ratio in Eq. (3) is very similar to that adopted in Ising models for medium size DNA's (100–600 base pairs) [8,9]. This confirms that this term is related to the conformational flexibility of the double and single strands, not described by a one-dimensional model. The internal term is related to the one-dimensional Hamiltonian, and then to nearest neighbor interactions. For long DNA's (large N), at temperatures in which θ_{int} is already close to 0, the part in Eq. (2) depending

on the internal partition functions goes as $e^{-\gamma N}$ for some positive γ, and thus $\delta \approx 0$ and $\theta_{\text{ext}} \approx 1$. This N dependence of the internal part can be seen, for example, in the case of homogeneous sequences with the transfer matrix method [16–18]. It is expected to be the same for heterogeneous sequences.

In very short chains like ours, it is not surprising that the specific sequence has some influence on the parameters p and q, while in medium chains some self-averaging effects should already take place. In fact, as shown in the caption to Fig. 1, we have found differences of about 25% in the parameters referring to different sequences. We are now working on a more extended set of melting curves for a properly chosen set of oligonucleotides, that can help in the attempt to find the relation between the specific sequence and the optimized parameters. Such a study will also regard the quantitative assessment of the sensitivity of the melting curves to θ_{ext} and to the fitted parameters p and q, as a function of N. As we noted, in the limit of large N this sensitivity should go to 0. At the end it would be possible to test the predictive power of this model and confront it with the predictions of purely thermodynamical calculations.

In a more extended paper, in preparation, we will show a more exhaustive comparison with experimental curves. We will also check if a simple analysis based on the number of occurrences of the different intrastrand nearest neighbor couples in the sequences is sufficient to obtain the parameters, similarly to what happens in the calculations of Gibbs free energy in short oligonucleotides [6,7].

ACKNOWLEDGMENTS

We are very grateful to F. Barone, M. Matzeu, F. Mazzei, and F. Pedone for providing the experimental melting curves and for illuminating discussions. We also want to thank M. Peyrard for interesting comments.

[1] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry* (Freeman, San Francisco, 1980), Vol. III.
[2] L. A. Marky, K. S. Blumenfeld, S. Kozlowski, and K. J. Breslauer, Biopolymers **22**, 1247 (1983).
[3] K. J. Breslauer, Methods Enzymol. **259**, 221 (1995).
[4] E. T. Kool, Annu. Rev. Biophys. Biomol. Struct. **25**, 1 (1996).
[5] M. Doctycz *et al.*, J. Biol. Chem. **270**, 8439 (1995).
[6] K. J. Breslauer, R. Franck, H. Blöcker, and L. A. Marky, Proc. Natl. Acad. Sci. USA **83**, 3746 (1986).
[7] J. SantaLucia, H. T. Allawi, and P. A. Seneviratne, Biochemistry **35**, 3555 (1996).
[8] D. Poland and H. R. Scheraga, *Theory of Helix-Coil Transition in Biopolymers* (Academic, New York, 1970).
[9] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
[10] T. Dauxois and M. Peyrard, Phys. Rev. E **51**, 4027 (1995).
[11] Y. Z. Chen and E. W. Prohofsky, Eur. Biophys. J. **25**, 9 (1996).
[12] L. V. Yakushevich, Studia Biophys. **121**, 201 (1987).
[13] L. V. Yakushevich, Phys. Lett. A **136**, 413 (1989).
[14] G. Gaeta, C. Reiss, M. Peyrard, and T. Dauxois, Riv. Nuovo Cimento **17**, 1 (1994).
[15] M. Salerno, Phys. Rev. A **44**, 5292 (1991).
[16] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
[17] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, 684 (1993).
[18] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, R44 (1993).
[19] D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).
[20] Yong-li Zhang, Wei-Mou Zheng, Ji-Xing Liu, and Y. Z. Chen, Phys. Rev. E **56**, 7100 (1997).
[21] These temperatures differ by about 40 K, and their exact values depend on the solvent conditions, especially the ionic strength of the solution. This means that in any effective model, dynamical or thermodynamical, the parameters should depend on the solvent conditions.
[22] We are considering non-self-complementary strands, i.e., $A \neq B$.
[23] To reproduce the experimental conditions, one should use isobaric-isothermal (P,T), rather than canonical isochoric-isothermal (V,T), partition functions. The practical ideality of the solution allows the use of the latter without appreciable differences.
[24] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon, Oxford, 1980).