Melting of DNA Oligomers: Dynamical Models and Comparison with Experimental Results

A. CAMPA¹ and A. GIANSAANTI²

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 INFN Unità Roma 1, Piazzale Aldo Moro, 2, 00185 Roma, Italy*

Accepted in final form 13 November 1998

Abstract. We compare experimental melting curves of short heterogeneous DNA oligomers with theoretical curves derived from statistical mechanics. Partition functions are computed with the one-dimensional Peyrard-Bishop (PB) Hamiltonian, already used in the study of the melting of long DNA chains. Working with short chains we take into account, in the computations, not only the breaking of the interstrand hydrogen bonds, but also the complete dissociation of the double helix into separate single strands. Since this dissociation equilibrium is of general relevance, independent of the particular microscopic model, we give some details of its treatment. We discuss how the non bonded three-dimensional interactions, not explicitly considered in the one-dimensional PB model, are taken into account through the treatment of the dissociation equilibrium. We also evaluate the relevance of the dissociation as a function of the chain length.

Key words: Dissociation equilibrium, melting, non-bonded interactions, nonlinear models, oligonucleotides, short duplex DNA

1. Introduction

The melting of DNA is the cooperative thermal disruption of hydrogen bonds between complementary bases in the double helix, detected by the UV absorption increment due to the unstacking of the separated bases. Although short DNA chains (oligonucleotides) are not naturally found in biological structures, they have a great relevance in many different biological techniques, such as: polymerase chain reaction (PCR), sequencing by hybridization, antigene targeting, Southern blotting [1–3]. All these techniques are of basic importance in modern molecular genetics. In many of these applications it would be useful to be able to predict the melting curves of short oligonucleotides of a prescribed sequence, or at least the melting temperature. Then it is very important to have a microscopic theoretical understanding of the melting process. After the pioneering work of Breslauer and Marky [4], there has been remarkable progress towards the prediction of the sequence-dependent melting temperature of short duplex DNA, based on nearest-neighbour thermodynamics [5, 6], and not on a microscopic model.

In this paper we present a theoretical approach to the interpretation and prediction of the melting curves of short DNA chains, based on a microscopic model and on statistical mechanics; as it will be shown, short means DNA chains with no more than few hundreds base pairs. The microscopic modeling of a molecular system, based on a Hamiltonian, requires an explicit expression for the molecular forces that regulate the structure and function of DNA. These kinds of models have been introduced mainly for the study of nonlinear energy localization and transduction along biopolymers [7], and, more recently, by Peyrard and Bishop (PB) in 1989 also for the study of DNA melting [8]. Confrontation with a set of experimental data is the only way to validate a model and to fix the values of the relevant parameters. With respect to a purely thermodynamical approach, the use of microscopic models has the advantage that a validated model can be used to study the dynamics of the system. In a short article we have already reproduced some experimental melting curves of short DNA duplexes via the PB model [9]; in this paper we present the details of the method of computation, comparison with further experimental curves, and an extended discussion of different topics relevant for our modelization.

In the following Section we discuss the problem of taking into account the equilibrium of dissociation of the duplex into single strands. In Section 3 we introduce the Peyrard-Bishop model and we show how to compute the local opening of the hydrogen bonds. In Section 4 we show our results and the comparison with several experimental curves; we also make an evaluation of the importance of the dissociation equilibrium, in the calculation of the melting curves, as a function of the chain length N . In the last Section we give some comments and conclusions.

2. The dissociation equilibrium

For long DNA fragments, when the fraction θ of intact hydrogen bonds between complementary bases goes practically from 1 to 0 at the melting transition, the two strands are not yet completely separated; the great majority of the bonds is disrupted and the DNA has denatured, but the few bonds still remaining prevent the two strands from going apart each other [10, 11]. Only at higher temperatures will there be a real separation. Therefore the double strand is always a single macromolecule through the transition, and, in a calculation based on a given model, one has to compute the fraction of intact or disrupted bonds.

The situation is more difficult for short fragments. In this case the processes of single bond disruption and strand dissociation tend to happen in the same temperature range. As a consequence, the value of θ is given by the product of two quantities: the fraction of strands that exist in the duplex form, that can be called θ_{ext} , and the fraction of intact bonds in the existing double strands, that can be called θ_{int} [10, 11]. This factorization $\theta = \theta_{int}\theta_{ext}$ is in principle valid in general; however, for long fragments, where at the melting temperature strand dissociation is still negligible, the part of the calculation concerning θ_{ext} can be neglected, as far

as reproduction of melting curves is concerned, since θ_{int} and thus θ go to 0 when θ_{ext} is still practically 1. On the contrary, for short DNA fragments both factors have to be taken into account. In Section 4 we will evaluate the relevance of θ_{ext} as a function of the chain length, and we will see that for DNAs longer than a few hundred base pairs it can be neglected. This can give a definition of shortness for DNA chains in the present context.

It is possible to envisage a statistical mechanics calculation for both θ_{int} and θ_{ext} . In this Section we will give some details of how this can be done for θ_{ext} , deferring to the next Section the calculation of θ_{int} .

The starting point is the thermodynamical relation governing the equilibrium between reacting species. Suppose there are in solution two types of single strands, A and B (we are considering non self-complementary strands), that can associate to give the duplex AB . We do not consider self-complementary strands simply because our experimental data refer to non self-complementary strands. However, we note that treating self-complementary strands would imply only minor differences in the formulas concerning the dissociation equilibrium (the probability of ‘hairpin’ formation is small for short sequences). At equilibrium the following equation holds for the chemical potentials of the three species:

$$\mu_{AB} - \mu_A - \mu_B = 0. \quad (1)$$

This relation can be expressed in terms of partition functions. For a system at the inverse temperature β with p different kinds of particles and with total Hamiltonian H , the canonical partition function Q

$$Q(N_1, \dots, N_p) = \left(\prod_{j=1}^p N_j! \right)^{-1} \int d\Gamma e^{-\beta H} \quad (2)$$

(where $d\Gamma$ means the volume element in the whole phase space of the system, and in Q we have made explicit only the dependence on the number of particles) and its relation to the Helmholtz free energy F , i.e. $F = -\frac{\ln Q}{\beta}$, allow, in principle, the evaluation of the chemical potentials $\mu_i = \frac{\partial F}{\partial N_i}$. Then, if we have a solution with N_A and N_B single strands of type A and B respectively, N_{AB} double strands, and N_s solvent molecules (although these will be of several different types we indicate them cumulatively since they will not enter our calculation), we have, e.g. for μ_A , the expression:

$$\mu_A = \frac{1}{\beta} \ln \left[\frac{Q(N_s, N_A, N_B, N_{AB})}{Q(N_s, N_A + 1, N_B, N_{AB})} \right] \quad (3)$$

and analogously for μ_B and μ_{AB} . Therefore, (1) can be rewritten as:

$$\frac{1}{\beta} \ln \left[\frac{Q(N_s, N_A + 1, N_B, N_{AB}) Q(N_s, N_A, N_B + 1, N_{AB})}{Q(N_s, N_A, N_B, N_{AB}) Q(N_s, N_A, N_B, N_{AB} + 1)} \right] = 0 \quad (4)$$

If in this expression we separate the momentum and the configuration parts, the former simplifies, and we obtain:

$$\frac{N_{AB}Z(N_s, N_A + 1, N_B, N_{AB})Z(N_s, N_A, N_B + 1, N_{AB})}{N_A N_B Z(N_s, N_A, N_B, N_{AB})Z(N_s, N_A, N_B, N_{AB} + 1)} = 1 \quad (5)$$

where with Z we denote the configurational partition function:

$$Z(N_s, N_A, N_B, N_{AB}) = \int d\Gamma e^{-\beta U}, \quad (6)$$

where $d\Gamma$ now denotes only a configuration volume element and U is the total potential. We can obtain a more manageable expression with the following considerations. Firstly, we want to study the typical experimental conditions, in which the solutions can be considered ideal, i.e., the concentrations are sufficiently low to ignore, in the computation of the partition functions, the intermolecular interaction between single strands, between duplexes, and between single strands and duplexes. Secondly, also in view of the use of simple dynamical models, the presence of the solvent molecules can be thought to be accounted for in an effective way, through a Hamiltonian describing the intramolecular interactions of single strands and duplexes. Then, it is not difficult to transform equation (5) into:

$$\frac{N_{AB}Z(A)Z(B)}{N_A N_B Z(AB)} = 1 \quad (7)$$

where $Z(A)$ is the configurational partition function of a single molecule A in the volume V of the system, and analogously for $Z(B)$ and $Z(AB)$. In the last expression, we have the constraints:

$$\begin{aligned} 2N_{AB} + N_A + N_B &= \text{constant} \equiv 2N_0 \\ \Delta N_A &= \Delta N_B = -\Delta N_{AB}. \end{aligned} \quad (8)$$

Now suppose to start with equal concentrations of A and B , i.e. $N_A = N_B$ (generally the experimental melting curves, and in particular those that we are going to show, are obtained in this way, with the duplexes formed by annealing equal concentrations of the complementary single strands). This equality is conserved because of the second constraint. Using also the first constraint, Equation (7) becomes a quadratic equation in $\theta_{ext} \equiv N_{AB}/N_0$, with solution:

$$\theta_{ext} = 1 + \delta - \sqrt{\delta^2 + 2\delta} \quad (9)$$

where the nonnegative variable δ is given by:

$$\delta = \frac{Z(A)Z(B)}{2N_0 Z(AB)}. \quad (10)$$

Equations (9) and (10) are the final expressions governing the dissociation equilibrium. Since θ_{ext} is a decreasing function of δ , and, on physical grounds, it should

be a non-increasing function of the temperature, this implies that δ should be an increasing function of the temperature.

The partition functions appearing in (10) have to include the statistical weights of all the possible configurations for the three different species A , B and AB , and therefore they should take into account all the interactions. Since we will use a simple model that describes only local bonded interactions, it will be necessary to include the other, non-bonded, interactions, in a phenomenological way, that will be described in Section 4.

A last remark. It should be noted that, in order to reproduce the experimental conditions, one should consider the Gibbs free energy, G , instead of the Helmholtz free energy, F ; therefore, instead of the NVT canonical ensemble, one should consider the NPT isothermal-isobaric ensemble. At the end one would arrive at Equation (7) with now $Z(A)$, $Z(B)$ and $Z(AB)$ denoting configurational isothermal-isobaric partition functions. However, the practical ideality of the solutions allows the use of the NVT ensemble without appreciable differences.

3. The model

Modeling of DNA melting was initially based on Ising-like models [10, 11], and more recently on an approach based on the modified self-consistent phonon approximation [12] (see also [13] and references therein). These methods allow only equilibrium estimates of the probability of bond disruption. However, there is also an interest in relaxation and kinetic phenomena, which are relevant for pharmacological applications [14], and in the study of nonlinear energy localization and transduction via soliton-like excitations. With a particular focus on the last problem, discrete nonlinear models of DNA (see, e.g., [15, 16], and for a review [7] and the contribution of G. Gaeta to this volume), have been introduced; sequence effects have been considered in [17] and [18]. These are microscopic models with a small number of degrees of freedom.

The authors of the PB model have shown, through statistical mechanics calculations and constant temperature molecular dynamics [8, 19, 20], applied to the case of a very long homogeneous DNA chain, that it can give a satisfactory melting curve, especially after the improvement introduced in [20]. The PB model has been successively applied to heterogeneous chains, either modeling the heterogeneity with a quenched disorder [21], or properly choosing basis sets of orthonormal functions for the kernels appearing in the expression of the partition function [22]. In all these papers, only very long chains were considered, and comparison with experimental data was not attempted. To our knowledge, we are the first to apply the PB model to the melting of short heterogeneous DNAs.

The PB model describes the double helix with two degrees of freedom per base pair, one for each nucleotide of the pair. The interactions considered are the transverse interaction of the two bases in a pair, connected by hydrogen bonds, and the nearest neighbour intrastrand (stacking) interaction of the bases. In the

Figure 1. Computed melting temperature vs. G-C content for the PB model applied to very long DNA chains. The computation has been performed through the transfer matrix method, and chains with small repeating blocks have been used (e.g, G-C fraction equal to 0.75 means repetition of quadruplets, each with three G-C pairs and one A-T pair). For large scale heterogeneity we expect a multiphasic behaviour and a less defined melting temperature.

calculations regarding melting, the relevant part of the potential energy depends on only one degree of freedom per base pair, and it is given by:

$$U = \sum_i \left\{ \frac{k}{2} [1 + \rho e^{-\alpha(y_{i+1} + y_i)}] (y_{i+1} - y_i)^2 + D_i (e^{-a_i y_i} - 1)^2 \right\} \quad (11)$$

where y_i is the distance between the i -th complementary bases minus their equilibrium separation. The parameters k , ρ and α refer to the anharmonic stacking interaction, while the interbase hydrogen bond is represented by a Morse potential, with depth D_i and width a_i . The ρ dependent nonlinear term was found to be relevant to give cooperativity to the melting process of long DNAs [20]. To model heterogeneous DNAs, we have inserted two different values of D_i , according to the two possible Watson-Crick base pairs: adenine-thymine (A-T) and guanine-cytosine (G-C). The former has two hydrogen bonds, while the latter has three. We have then chosen a depth for the G-C Morse potential 1.5 times that for the A-T Morse potential. For a given set of values, the melting temperature of homogeneous DNA can be deduced with the technique of the transfer matrix method [8, 19, 20], and the values have been adjusted to reproduce the experimentally observed melting temperature of long homogeneous DNA, all A-T and all G-C, in the most usual solvent conditions [11, 23]. Once fixed in this way the melting temperatures for these two cases, we have checked that for long heterogeneous DNA the melting temperature is linearly dependent, between the two values, on the G-C percentage (Figure 1). For this last computation we have used a simple extension of the transfer matrix method, in which a small repeating block (for example, for a G-C fraction of 0.75 we have repeating quadruplets made of three G-C pairs and one A-T pair) substitutes the single base pair of homogeneous chains, as the repeating element that allows to express the partition function as the product of identical matrices.

The complete set of parameters that we have chosen is: $k = 0.025 \text{ eV/\AA}^2$, $\rho = 2$, $\alpha = 0.35 \text{ \AA}^{-1}$, $D_{AT} = 0.05 \text{ eV}$, $D_{GC} = 0.075 \text{ eV}$, $a_{AT} = 4.2 \text{ \AA}^{-1}$, $a_{GC} = 6.9 \text{ \AA}^{-1}$. The set of values given above has been used for all the different oligonucleotides considered.

With the potential (11) we can compute the fraction θ_{int} of intact hydrogen bonds in existing duplexes. First we have 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 non trivial problem. We have adopted the following strategy. The i -th bond is considered disrupted if the value of y_i is larger than a chosen threshold y_0 . We have therefore 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:

$$\theta_{int} = \frac{1}{N} \sum_{i=1}^N \langle \vartheta(y_0 - y_i) \rangle \quad (12)$$

where $\vartheta(y)$ is the Heaviside step function and the canonical average $\langle \cdot \rangle$ is defined considering only the double strand configurations. We have chosen a value of $2 \text{ \AA}$ 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 [12], the computations needed for the canonical averages are readily reduced to the multiplication of finite matrices, since the potential (11) couples only nearest neighbours, and are easily performed by computer programs optimized for numerical matrix manipulation.

4. Results

In the first subsection we show the comparison of our calculations with experimental melting curves, and we show how to implement the expressions for the dissociation equilibrium. In the second subsection we will make an evaluation of the relevance of the dissociation equilibrium, as far as melting curves are concerned, as a function of the length N of the chain.

4.1 EXPERIMENTAL AND COMPUTED MELTING CURVES

We have at our disposal several melting curves of oligonucleotides that have obtained in the Physics Laboratory of our Institute. To reproduce them we have computed θ_{int} as in equation (12) and then we have used equations (9) and (10) for θ_{ext} . However, as noted at the end of Section 2, in order to include all the relevant factors determining the partition functions ratio given in (10), it is necessary to take into account in a phenomenological way the interactions not included in the simple one-dimensional model. To this purpose, we introduce the separation of the partition functions in an internal and an external part [10, 11], to transform (10) in:

$$\delta = \frac{Z_{int}(A)Z_{int}(B)}{a_{av}Z_{int}(AB)} \frac{a_{av}Z_{ext}(A)Z_{ext}(B)}{2N_0Z_{ext}(AB)}, \quad (13)$$

where a_{av} has been inserted for dimensional reasons, and its meaning 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 PB model (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). In (13) $a_{av} = \sqrt{a_{AT}a_{GC}}$ makes separately dimensionless both fractions, that therefore can not 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} , but in the external functions Z_{ext} . The necessity to include in Z_{ext} also what is not described by the one-dimensional model has already been considered in Ising models [10,11]. This factor is very likely related to non bonded interactions and therefore to three-dimensional movements (like, for example, the winding of the strands). In analogy to what has been proposed for the Ising models [10,11] on the basis of the partition functions of rigid bodies [24], we make the following choice:

$$\frac{a_{av}Z_{ext}(A)Z_{ext}(B)}{2N_0Z_{ext}(AB)} = \frac{n^*}{n_0}e^{-p\theta_{int}+q} \quad (14)$$

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 μ M, a usual concentration in experiments). We want to remark the following. If we would be able to describe the single and the double strands in a complete way, with an all-atom model and with all the interactions included, then the external partition functions would reduce to the translation of the whole macromolecule, and the left hand side of (14) would only be proportional to the inverse of the concentration of the solution. In fact, in an all-atom model we would separate the degrees of freedom of the molecule in the center of mass degrees of freedom and the rest; the internal potential energy would depend only on the latter, and integration on these would give the internal partition function. Afterwards, the only remaining integration, defining the external partition function, would be that on the center of mass, related to a global translation. Therefore a complete model would allow the calculation of θ_{ext} in a direct way; on the contrary, we expect that the calculation of θ_{int} , related to local interactions, is already well performed by the one-dimensional model (which is equivalent to say that for long DNAs, where the dissociation equilibrium can be neglected, the one-dimensional model already describes well the melting transition).

We show in Figures 2 to 5 the comparison between the experimental and the theoretical melting curves for six different oligonucleotides: three with 21 base pairs, two with 27 base pairs and one with 42 base pairs. The strand sequences of our oligomers are given by:

Figure 2. Experimental melting profiles (dotted line) and theoretical results (solid line) for the DNA oligomer corresponding respectively to s_1 (a) and s_2 (b). The buffer system used for these and all the other oligomers was composed by 10 mM sodium phosphate, 0.1 mM sodium EDTA, 200 mM sodium chloride, adjusted to pH 6.7. The fitted parameters have the values: $p = 111.95$ and $q = 106.22$ for (a), while $p = 94.00$ and $q = 88.52$ for (b).

s_3 : 5'CTTCTTATTCTTATTGTTTCGTCTTCTC_{3'}
 s_4 : 5'CGATCTCTCTCTCTTTTTTTTTTTTAGC_{3'}
 s_5 : 5'GGGCAAAAACGGGCGGCCAAAATTTTG
 CCGCCCATCGATGGG_{3'}
 s_6 : 5'GTGTTAACGTGAGTATAGCGT_{3'}

and by the respective complementary strands. We have the experimental curves for oligomer s_6 at two different concentrations. The single strand concentrations are: s_1 : 1.7 μM , s_2 : 2.2 μM , s_3 : 2.4 μM , s_4 : 1.7 μM , s_5 : 1.4 μM , s_6 : 3.1 and 120 μM . In the figures we have plotted for convenience $\phi \equiv 1 - \theta$. The experimental data concern also quite large temperature ranges outside the transition region. The experimental data concerning the oligomer s_6 have been published in [25], those concerning s_4 are in [26], while the other data are a private communication. We have concentrated the computations in this region (outside we would get $\phi = 0$

Figure 3. Experimental melting profiles (dotted line) and theoretical results (solid line) for the DNA oligomer corresponding respectively to s_3 (a) and s_4 (b). The fitted parameters have the values: $p = 106.88$ and $q = 96.57$ for (a), while $p = 87.64$ and $q = 82.35$ for (b).

Figure 4. Experimental melting profiles (dotted line) and theoretical results (solid line) for the DNA oligomer corresponding to s_5 . The fitted parameters have the values: $p = 54.56$ and $q = 51.69$.

Figure 5. Experimental melting profiles (dotted line at the lower concentration and dash-dotted line at the higher concentration) and theoretical results (solid lines) for the DNA oligomer corresponding to s_6 . The fitted parameters have the values: $p = 89.78$ and $q = 84.30$.

and $\phi = 1$ at low and high temperatures, respectively). This can be noted in some of the graphs, where the theoretical curves do not extend for all the temperature axis. Our computations, as said above, take into account both factors: θ_{int} and θ_{ext} . To assess quantitatively the relevance of θ_{ext} we have checked that at the complete melting, when θ becomes practically zero (i.e., $\phi = 1$), the values of θ_{int} are still: 0.85 for s_1 , 0.84 for s_2 , 0.81 for s_3 , 0.82 for s_4 , 0.80 for s_5 , 0.83 for s_6 at the lower concentration, and 0.78 for s_6 at the higher concentration. This is consistent with the Ising-like computation shown by Wartell and Benight for a 95 base pairs DNA, where θ_{int} at complete melting is still 0.63 (see Fig. 8 of Ref. [11]). Note that a 40 fold concentration increase for s_6 yields an increase of only 5 degrees in the melting temperature (a logarithmic dependence on the concentration is expected [27]). We would like to stress that in the case s_6 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 equation (14), 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 degrees. This fact indicates that the concentration dependence of the left hand side of equation (14) is described by the preexponential factor, while the parameters p and q are related to the molecular three-dimensional dynamics. In the last Section we will try to give an explanation of the possible source of the difference, among the various oligomers, of the values of p and q .

From the above analysis we have to deduce that for lengths of few tens of base pairs the dissociation equilibrium is the more relevant factor for the melting transition. To have a confirmation of this, we have performed again the computation with a the nonlinear parameter ρ of the stacking interaction set equal to 0 (and adjusting, as before, the other parameters in order to have proper melting temperatures for long DNA's). It was shown in [20] that the main effect of the

parameter ρ is to narrow the melting transition in long chains. Therefore we expect that for short chains this term has a minor influence on the melting curves, since the melting happens in a temperature range where the “internal” transition, given by θ_{int} , has not yet been reached. In fact, without showing graphs analogous to those in Figures 2 to 5, we can say that also with $\rho = 0$ we found a fair agreement between experimental and theoretical curves. We have also checked that a smaller threshold y_0 (e.g., 1 Å) does not change appreciably θ_{int} at the experimental melting temperatures.

We conclude this subsection with a proposal to evaluate the energies involved in the interactions governing the dissociation equilibrium. In the temperature range in which the melting of our oligomers takes place, θ_{int} decreases roughly linearly with the temperature. It is therefore possible to substitute the exponential term appearing in (14) with:

$$be^{c\Delta T} \quad (15)$$

where now the parameters to fit are b and c , and the temperature variation ΔT in the transition region can be measured, for example, with respect to the temperature at the beginning of the transition. We have checked that, in all cases considered, the exponential term in (14) or in (15) increases about 10^3 times from the beginning to the end of the transition. Then, if we indicate with $(\Delta T)_0$ the temperature range of the transition, we have $\exp\{c(\Delta T)_0\} \approx 10^3$. We have already noticed that Z_{ext} is presumably connected to non bonded interactions and three-dimensional movements, which, however, are related to the variation of θ_{int} [10,11]. In the same spirit of the self-consistent phonon method, we make the hypothesis that we can model in an effective way the potential related to this movements, in the temperature range of the transition, with a temperature dependent potential with which we can compute Z_{ext} . We also make the further simplifying but reasonable assumption that it is sufficient to consider this kind of effective interaction only in the double strands. More precisely, we compute the temperature variation of the external partition functions ratio appearing in the left hand side of (14) evaluating the variation of $Z_{ext}(AB)$, supposing that this, in the temperature range of the transition, is determined by a potential of the form:

$$U_{nb}(T) = U_{nb}(T_1) + \frac{\Delta U_{nb}}{(\Delta T)_0}(T - T_1) \quad (16)$$

where the subscript nb stands for ‘non-bonded’, and T_1 and T_2 are the temperatures of the beginning and the end of the transition, with $(\Delta T)_0 = T_2 - T_1$. With this modelization, it is not difficult to see that the parameter c in (15) is related to ΔU_{nb} . From the estimate of the variation of (15) through the transition we then derive an estimate for ΔU_{nb} of about 0.2 eV, a quite reasonable value. We want to stress that with this crude procedure we do not want to propose a modelization of the non-bonded interactions, but only evaluate a typical value of the energies involved.

Figure 6. Upper curve: melting temperature, as a function of N , as computed in the PB model neglecting the dissociation equilibrium, i.e., supposing that $\theta_{ext} = 1$ and therefore $\theta = \theta_{int}$. Lower curve: melting temperature, as computed through the empirical law $1/T_m = 1/T_m^\infty + B/N$, derived by fitting experimental data on synthetic oligomers (see [27], chapter 23, where also a typical value for B can be found).

4.2 DISSOCIATION VS. N

In this subsection we want to make an evaluation of the importance of the dissociation equilibrium as a function of the chain length N . To fix the ideas, we consider the case of homogeneous G-C chains; our conclusions should not depend on this choice.

In Figure 6 we compare two curves. The first curve gives the melting temperature, as a function of N , as computed in the PB model neglecting the dissociation equilibrium, i.e., supposing that $\theta_{ext} = 1$ and therefore $\theta = \theta_{int}$. The second curve gives the melting temperature, as computed through an empirical law (see caption to the figure) derived by fitting experimental data on synthetic oligomers (see [27], chapter 23); while we can not say that this curve is extremely exact in giving the melting temperature, nevertheless for our aims here it is sufficient. From the comparison between the two curves we can see that for lengths greater than few hundreds base pairs the melting is determined by θ_{int} , while for shorter lengths the dissociation equilibrium is necessary to have a proper evaluation of the melting temperature. This confirms what we have anticipated at the beginning of Section 2.

Looking at (9) and (10) and at the immediately following remark, we can say that, when N is large enough, before δ becomes of order 1, increasing the temperature, and thus θ_{ext} begins to decrease from 1, the “internal” melting governed by θ_{int} takes place.

5. Conclusions

We would like to note that the modelization of the external partition functions ratio as in equation (14) is very similar to that adopted in Ising models for medium sized DNA’s (100–600 base pairs) [10, 11]. This confirms that this term is related to the three-dimensional movements, like 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 local (nearest neighbour) interactions. In this respect we note that equation (14) tells implicitly that there is an interaction between local and three-dimensional movements. We expect that the strength of this interaction depends on both the length of the chain and the sequence. This should explain the differences among the values of p and q in the different oligomers. The chains for which we have made our comparisons with experimental melting curves have few tens of base pairs. For longer chains (but still in the range where dissociation equilibrium is important) it could be argued that the values of p and q tend to be more universal, depending only on the composition (i.e., G-C percentage) and not on the sequence.

In subsection 4.1 we proposed a way to evaluate the energies involved in the dissociation equilibrium. The underlying idea was to relate the local bonded interactions with the non-bonded interactions. The logical consequence of this approach would be to build a more complete model in which this relation is considered. A first step in this direction is the model proposed by Barbi et al. in another paper in this volume (paper which is however devoted to breather propagation). It is an improvement of the PB model, with the insertion of the angular displacement between neighbouring base pairs, in order to take into account torsional degrees of freedom and their relation with the stretching of the hydrogen bonds. It could be interesting to study the effect of this enrichment of the model on the computation of the dissociation equilibrium of short DNAs.

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, and to M. Peyrard for interesting comments. It is also a pleasure to warmly thank L. V. Yakushevich and all the local organizing team of the Institute of Cell Biophysics for the enjoyable and scientifically stimulating atmosphere in Pushchino.

References

1. Uhlmann, E. and Peyman, A.: Antisense oligonucleotides: a new therapeutic principle, *Chem. Rev.* **90** (1990), 543–584.
2. Doktycz, M.J., Morris, M.D., Dormandy, S.J., Beattie, K.L. and Jacobson, K.B.: Optical melting of 128 octamer DNA duplexes, *J. Biol. Chem.* **270** (1995), 8439–8445.
3. SantaLucia, J. Jr., Allawi, H.T. and Seneviratne, P.A.: Improved nearest-neighbor parameters for predicting DNA duplex stability, *Biochemistry* **35** (1996), 3555–3562.
4. Marky, L.A. and Breslauer, K.J.: Calorimetric determination of base-stacking enthalpies in double-helical DNA molecules, *Biopolymers* **21** (1982), 2185–2194.
5. Owczarzy, R., Vallone, P.M., Gallo, F.J., Paner, T.M., Lane, M.J. and Benight, A.S.: Predicting sequence-dependent melting stability of short duplex DNA oligomers, *Biopolymers* **44** (1997), 217–239.

6. SantaLucia, J. Jr.: A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc. Natl. Acad. Sci. USA* **95** (1998), 1460–1465.
7. Gaeta, G., Reiss, C., Peyrard, M. and Dauxois, T.: Simple models of non-linear DNA dynamics, *Rivista del Nuovo Cimento* **17** (1994), 1–48.
8. Peyrard, M. and Bishop, A.R.: Statistical mechanics of a nonlinear model for DNA denaturation, *Phys. Rev. Lett.* **62** (1989), 2755–2758.
9. Campa, A. and Giansanti, A.: Experimental tests of the Peyrard-Bishop model applied to the melting of very short DNA chains, *Phys. Rev. E* **58** (1998), 3585–3588.
10. Poland, D. and Scheraga, H.R.: *Theory of helix-coil transition in biopolymers*. New York: Academic Press (1970).
11. Wartell, R.M. and Benight, A.S.: Thermal denaturation of DNA molecules: a comparison of theory with experiment, *Physics Reports* **126** (1985), 67–107.
12. Dauxois, T. and Peyrard, M.: Entropy-driven transition in a one-dimensional system, *Phys. Rev. E* **51** (1995), 4027–4040.
13. Chen, Y.Z. and Prohofsky, E.W.: Sequence and temperature effect on hydrogen bond disruption in DNA determined by a statistical analysis, *Eur. Biophys. J.* **25** (1996), 9–18.
14. Kool, E.T.: Circular oligonucleotides: new concepts in oligonucleotide design, *Annu. Rev. Biophys. Biomol. Struct.* **25** (1996), 1–28.
15. Yakushevich, L.V.: The effect of damping, external fields and inhomogeneity on the nonlinear dynamics of biopolymers, *Studia Biophys.* **121** (1987), 201–207.
16. Yakushevich, L.V.: Nonlinear DNA dynamics: a new model, *Phys. Lett. A* **136** (1989), 413–417.
17. Salerno, M.: Discrete model for DNA-promoter dynamics, *Phys. Rev. A* **44** (1991), 5292–5297.
18. Peyrard, M. and Dauxois T.: Role of the base sequence on the dynamics of DNA melting, in R. Gruber and M. Tomassini (eds.), *Proc. 6th Joint EPS-APS International Conference on Physics Computing*, Geneva: European Physical Society, 1994.
19. Dauxois, T., Peyrard, M. and Bishop, A.R.: Dynamics and thermodynamics of a nonlinear model for DNA denaturation, *Phys. Rev. E* **47** (1993), 684–695.
20. Dauxois, T., Peyrard, M. and Bishop, A.R.: Entropy-driven DNA denaturation, *Phys. Rev. E* **47** (1993), R44–47.
21. Cule, D. and Hwa, T.: Denaturation of heterogeneous DNA, *Phys. Rev. Lett.* **79** (1997), 2375–2378.
22. Zhang, Y.L., Zheng, W.M., Liu, J.X. and Chen Y.C.: Theory of DNA melting based on the Peyrard-Bishop model, *Phys. Rev. E* **56** (1997), 7100–7115.
23. 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.
24. Landau, L.D. and Lifshitz, E.M.: *Statistical Physics*. Oxford: Pergamon Press, 1980.
25. Bonincontro, A., Matzeu, M., Mazzei, F., Minoprio, A. and Pedone, F.: Influence of defects on the electrophoretic, thermodynamic and dielectric properties of a 21 base pair DNA in solution, *Biochim. Biophys. Acta* **1171** (1993), 288–294.
26. Barone, F., Matzeu, M., Mazzei, F. and Pedone, F.: Stability and conformation of triplex oligonucleotides formed on defective duplexes, preprint, 1998.
27. Cantor, R.C. and Schimmel, P.R.: *Biophysical Chemistry*. San Francisco: W. H. Freeman and Company (1980). Vol. III.