# Study of the thermal stability of short DNA duplexes via the Peyrard-Bishop model

A. Campa[1] and A. Giansanti[2]

[1]Physics Laboratory, Istituto Superiore di Sanità
and INFN Sezione Sanità
Viale Regina Elena, 299 - 00161 Roma, Italy
e-mail: campa@axiss.iss.infn.it Fax: +39-6-4462872

[2]Physics Department, University of Rome "La Sapienza"
P.le Aldo Moro, 2 - 00185 Roma, Italy
e-mail: giansan@physig.sci.uniroma1.it Fax: +39-6-4463158

**Abstract**

The melting curves of short sequences of DNA is solution are calculated on the basis of statistical thermodynamics. The computation of the partition function is based on the Peyrard-Bishop hamiltonian, which has already been adopted in the theoretical description of the melting of long DNA chains. We have adapted this model to short sequences, where it is necessary to consider also the complete dissociation of the two strands, contrary to the case of long sequences. We show a comparison with some experimental data.

## 1 Introduction

Thermodynamic characterization and molecular understanding of DNA melting is basic in almost all fields of pure and applied molecular biology. It is also a classic but still challenging problem in statistical physics [1,2].

The thermal disruption of the interbase hydrogen bonds in heterogeneous DNA in solution appears to be a complex multistep process, very much affected by the ionicity, pH and composition of the solvent, by the length of the chain and by the particular sequence of the bases in the paired strains.

UV absorbance of DNA solutions, monitored as a function of the temperature, has been the main experimental technique to follow the thermal unstacking of the bases [3].

The systematic approach to an understanding of the thermal stability of complex DNA structures has been based on the experimental study of oligonucleotides of defined length and sequence. Using a large amount of thermodynamic data on short oligonucleotides, and through a simple analysis of all possible nearest neighbours stacking interactions [4,5] the melting temperatures of different short sequences has been predicted with good approximation. Despite this success, this purely thermodynamical approach does not give any insight of the melting process at the molecular level. On the other hand molecular dynamics (MD) computer simulations of all atoms models of the oligonucleotides still appear to be unpractical because of the large simulation time that would be required to study the melting process. There is then a need for tractable atomic-level models. The study of DNA dynamics via continous and discrete nonlinear models is a well established subject [6,7] essentially devoted to the investigation of the the different aspects of energy localization and transduction. In 1993 M. Peyrard and collaborators made an extensive study of the thermal stability of a nonlinear model of homogeneous long chains of DNA based on statistical thermodynamics and on constant temperature MD simulations [8,9]. Recently the Peyrard-Bishop hamiltonian has been used to investigate the multi-step behaviour in the melting of long heterogeneous (disordered) DNA chains [10].

In the present project we use the Peyrard-Bishop hamiltonian to compute the observed melting profile of oligonucleotides of fixed sequences. The importance of taking into account sequence effects in the realistic nonlinear modelling of the DNA behaviour was recently stressed by M. Salerno [11].

In the following we show the model, we sketch the basic points of our computations, and then we show the comparison with some experimental data.

## 2 The model

The Peyrard-Bishop model [8,9] describes the DNA double strand 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 (modelled by a Morse potential), and the nearest neighbour intrastrand interaction of the bases. Calling with $v_i$ and $w_i$ the displacements of the bases of the $i$-th pair, the potential is best expressed after a change of variable; for each $i$, $v_i$ and $w_i$ are linearly tranformed such that the new variables describe the distance $y_i$ between the $i$-th complementary bases, and their average displacement $x_i$. The potential is given by:

$$\begin{aligned} U &= \sum_i \Big\{ \frac{k}{2}(x_{i+1}-x_i)^2 \\ &+ \frac{k}{2}\left[1+\rho e^{-\alpha(y_{i+1}+y_i)}\right](y_{i+1}-y_i)^2 + D_i\Big(e^{-a_i y_i}-1\Big)^2\Big\} \end{aligned} \qquad (1)$$

where the parameters $k$, $\rho$ and $\alpha$ refer to the stacking interaction, while the bond of complementary bases is represented by a Morse potential, with depth $D_i$ and width $a_i$. In refs. [8,9] there is only a single parameter $D$ because of homogeneity. The stacking interaction, that in the first attempts [8] was purely harmonic ($\rho = 0$), decreases when the complementary bases get farther ($\rho$ positive) [9].

To model heterogeneous duplexes, we have inserted two different values of $D_i$, according to the two possible pairs (A-T and G-C). We have also changed somewhat the values of the parameters from those proposed in ref. [9], following a procedure that will be explained at the beginning of the results Section.

## 3 The method

The hypercromicity observed in absorption experiments when temperature is raised, is due to the disruption of the Watson-Crick hydrogen bonds between complementary bases. This phenomenon happens in a rather narrow temperature range; the temperature at which one can deduce, from the experimental curve, that half of the bonds are disrupted, is called the melting

temperature. For long DNA fragments the fraction $\theta$ of intact bonds goes practically from 1 to 0 increasing the temperature through this melting transition. This does not mean that the two strands are completely separated; the DNA has denaturated since the great majority of the bonds is disrupted, but the few bonds still remaining prevent the two strands from going apart each other. Only at higher temperatures will we have a real separation. Therefore in a calculation one can consider the double strand always as a single macromolecule, and compute, as a function of the temperature, 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. Thus, computing the fraction of intact bonds in a double strand is not sufficient to have $\theta$. Calling this fraction $\theta_{int}$, it is also necessary to compute the fraction of strands that exist in the duplex form, and that can be called $\theta_{ext}$. Then one has $\theta = \theta_{int}\theta_{ext}$ [1,2]. According to what we have said, for long fragments $\theta_{int}$ and thus $\theta$ go to 0 when $\theta_{ext}$ is still practically 1. On the contrary, for short DNA fragments an essential part of the calculation is that regarding strand dissociation. This is governed by the laws of chemical equilibrium between different species.

We have made a statistical mechanics computation, in which partition functions have been used to obtain both $\theta_{int}$ and $\theta_{ext}$. Let us first consider $\theta_{ext}$. Its evaluation requires, as noticed, to consider the chemical equilibrium between dissociated single strands and double strands associated in duplexes. We suppose to have in solution two types of single strands, $A$ and $B$ (we are considering non self-complementary strands), that can associate to give the duplex $AB$. Therefore at any temperature we have an equilibrium for the reaction:

$$A + B \leftrightarrow AB \tag{2}$$

At equilibrium the following equation holds for the chemical potentials of the three species:

$$\mu_{AB} - \mu_A - \mu_B = 0. \tag{3}$$

We now use in this equation the definition of the chemical potential as a derivative of a thermodynamic potential and the relation of this to a statistic sum. We therefore obtain an equation involving the appropriate partition functions of the systems, i. e. of the duplex $AB$ and of the single strands $A$ and $B$. For the evaluation of these partition functions we use the model

of DNA described above. The model is simply adapted to the description of single strands: only a harmonic stacking interaction remains, which is weaker than in the duplex, since now the term involving $\rho$ is 0. We insert in our computations two parameters that describe the overall conformation of the strands (that can not be included in simple one-dimensional models). The technical details of this procedure will be given elsewhere. Here it suffices to say that at the end we get an equation for $\theta_{ext}$ as a function of the temperature and of the concentration.

For the computation of $\theta_{int}$, i. e., of the fraction of intact bonds in double strands, we have used a criterium in which a threshold in the proper variable determines if a bond is broken or not. The first problem to face is to have a proper separation of the configurations describing a double strand on one hand, and separated 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 $\theta_{int}$ for an $N$ base pair duplex by:

$$\theta_{int} = \frac{1}{N} \sum_{i=1}^{N} < \vartheta(y_0 - y_i) > \tag{4}$$

where $\vartheta(y)$ is the Heavyside step function and the canonical average $< \cdot >$ is defined considering only the double strand configurations. The numerical computation of $\theta_{int}$ has been performed through a Matlab program, that is very effective for the matrix multiplications involved in the partition functions of nearest neighbour potential energies as (1).

# 4 Results

For the choice of the values of the parameters in the potential energy (1) we have adopted the following strategy, that has resulted in values somewhat different from those proposed in [9]. We have chosen two different values of $D_i$, one describing A-T base pairs, and the other G-C base pairs, and we have taken the G-C value 1.5 times the A-T value, according to the number of hydrogen bonds involved in the respective Watson-Crick bond. Then we have computed the melting temperature of long homogeneous DNA chains,

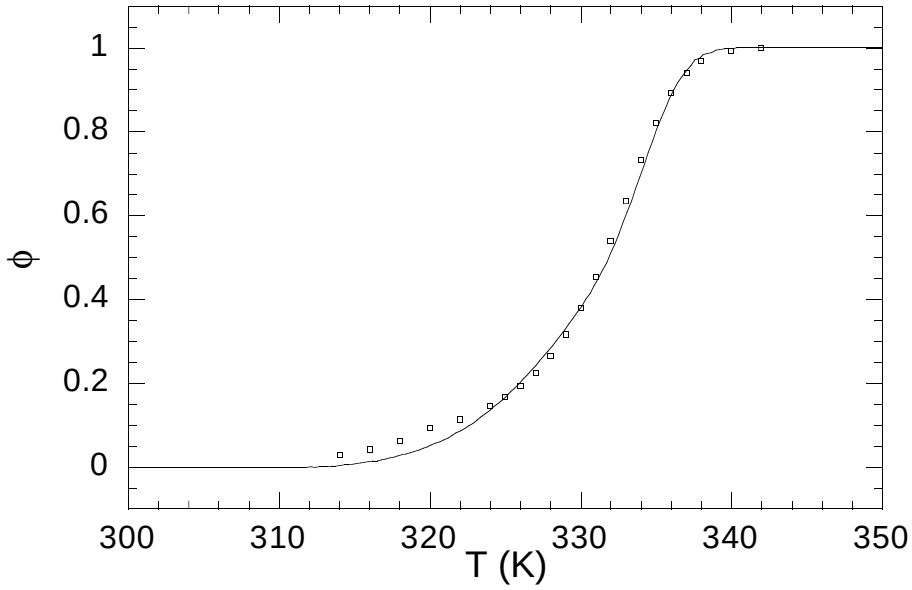


Figure 1: Experimental melting profiles (full curve) and theoretical results (open squares) for a 27 base DNA duplex. We have plotted the value of $\phi = 1 - \theta$. The parameter values used in the potential (1) are: $k = 0.025eV/\mathring{A}^2$, $\rho = 2$, $\alpha = 0.35\mathring{A}^{-1}$, $D_{AT} = 0.05eV$, $D_{GC} = 0.075eV$, $a_{AT} = 4.2\mathring{A}^{-1}$, $a_{GC} = 6.9\mathring{A}^{-1}$. The threshold value $y_0$ is $2\mathring{A}$.

for the case of only A-T and of only G-C chains, using the tranfer matrix method. We have then adapted the remaining parameters of the model in order to reproduced the experimentally observed melting temperature of long homogeneous DNA. These temperatures differ by about $40K$, and their exact values depend on the solvent conditions, especially the ionic strength of the solution (this means that in any effective model, in which the solvent molecules are not explicitly considered, the parameters should depend on the solvent conditions; we are then considering given conditions). Our values are mentioned in Figure 1 caption.

We show some comparison with experimental data. In Figure 1 we show the experimental and theoretical melting profiles of a 27 DNA duplex, while in Figure 2 we show the same for a 21 duplex of different composition. The two experimental curves have been obtained in the same solvent conditions. The agreement between the experimental and the theoretical curves is good,

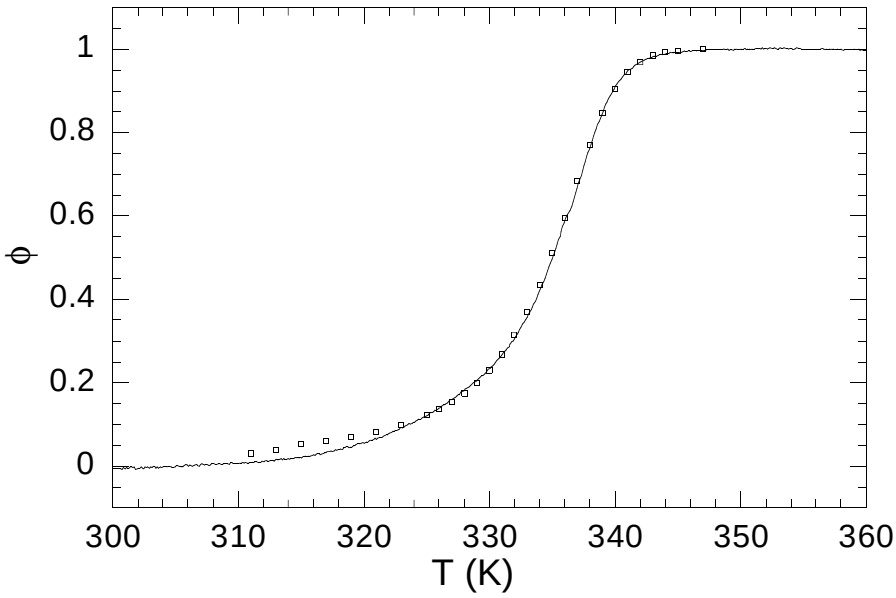


Figure 2: Same as Figure 1, but for a 21 base DNA duplex.

although there are some differences in the premelting part of the profile.

# 5 Conclusions

In this communication we have shown that it is feasible to compute the equilibrium melting profile of DNA oligonucleotides using in the calculation a simple nonlinear model for the atomic dynamics. With further work on a set of melting curves from a properly chosen set of oligonucleotide sequences we are setting down the method for the choice of the best parameters in the computation. The optimized parameters could be used to predict the melting profile of oligonucleotides not included in the reference set; the predictive power of our approach will be confronted with that of purely thermodynamical calculations.

A different theoretical approach to the same problem has been adopted by Prohofsky and collaborators [12]. In this approach a modified self-consistent phonon approximation is used to compute a statistical estimate of the probability of disruption of individual hydrogen bonds in a DNA base pair as a

function of temperature. The particular form chosen for the effective hamiltonian in these computations does not seem useful for further realistic dynamical studies at the molecular level.

## Acknowledgements

We are very grateful to F. Barone, M. Matzeu, F. Mazzei and F. Pedone for providing the experimental melting curves and for illuminating discussions.

## References

[1] D. Poland and H. R. Scheraga, *Theory of helix-coil transition in biopolymers*, (Acad. Press, New York, 1970).

[2] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).

[3] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry*, (W. H. Freeman and Company, San Francisco, 1980). Vol. III.

[4] K. J. Breslauer, Methods Enzymol. **259**, 221 (1995).

[5] J. SantaLucia, H. T. Allawi and P. A. Seneviratne, Biochemistry **35**, 3555 (1996).

[6] G. Gaeta et al., Rivista del Nuovo Cimento **17**, 1 (1994).

[7] L. V. Yakushevich, Studia Biophys. **121**, 201 (1987); Phys. Lett. **A 136**, 413 (1989).

[8] T. Dauxois, M. Peyrard and A. R. Bishop, Phys. Rev. **E 47**, 684 (1993).

[9] T. Dauxois, M. Peyrard and A. R. Bishop, Phys. Rev. **E 47**, R44 (1993).

[10] D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).

[11] M. Salerno, Phys. Rev. **A 44**, 5292 (1991).

[12] Y. Z. Chen and E. W. Prohofsky, Eur. Biophys. J. **25**, 9 (1996).