

Theoretical basis for the study of the effect of base composition on DNA melting

L. Dagdug

*Universidad Autónoma Metropolitana Iztapalapa,
Apartado Postal 55-534, 09340 México D.F. México*

L. Young

*Mathematical and Statistical Computing Laboratory,
National Institutes of Health, Bethesda, MD, 20892, USA*

Recibido el 22 de enero de 2004; aceptado el 14 de junio de 2004

We extend the ideas used to describe the glass transition in strong glasses employing the stochastic matrix method, giving a theoretical framework for the study of the configurational changes and the melting temperature of DNA. Our theoretical model enables a systematic study of the melting transition and the melting temperature dependence on the sequence differences in vertical stacking. Taking into account the fractional composition in a single strand, exact analytic results are given for the fraction of bonds intact and denatured at a particular temperature. This method is applicable to long DNA as well as RNA.

Keywords: DNA; melting temperature; stochastic matrix method.

Se implementan las ideas utilizadas para describir la transición vítrea en vidrios fuertes utilizando el método de la matriz estocástica, dando un marco teórico para el estudio de los cambios configuracionales y la temperatura de desnaturalización del ADN. Nuestro modelo teórico nos permite hacer un estudio sistemático de la temperatura de desnaturalización y de la dependencia de esta temperatura con respecto a la diferencia en la secuencia del ADN. Tomando en cuenta la composición en una cadena, se obtienen resultados analíticos para la fracción de cadena intacta y de la fracción denaturada a una temperatura en particular. El método es aplicable a cadenas largas de ADN como de RNA.

Descriptores: ADN; temperaturas de desnaturalización; método de la matriz estocástica.

PACS: 87.10+e

1. Introduction

It is well known that local denaturation of DNA is involved in the dynamics of DNA transcription, so it is interesting to investigate the denaturation of the double helix as a preliminary step to understanding it. Furthermore, the knowledge of the sequence dependence of DNA melting is important to understand the details of DNA replication, mutation, and repair. Accurate prediction of DNA thermal denaturation is also important for several biomolecular techniques including the PCR [1], followed by hybridization [2], antigen targeting [3], and Southern blotting [4]. In these techniques, the choice of a non optimal sequence or temperature can lead to an amplification or detection of a wrong sequence [5].

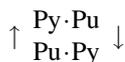
One of the ways to learn about the structure of macromolecules in solution is to observe structural changes, too. The ordered form of a nucleic acid is only marginally stable against temperature increase, so that most samples show a drastic alteration in structure within the convenient limits of 0°-100°C. Many physical properties are changed in the process, and the nature of these changes, and characteristics of the transformation provide fertile ground for physical studies [6].

The most common method of following the denaturation of DNA is the profile of ultraviolet absorbance against temperature, called the melting curve. An important quantity is the characteristic transition temperature (T_m). T_m is defined as the temperature at which half of the strands are in the double-helical state, and the other half are in the

“random-coil” state. A DNA melting curve is generally a two-dimensional plot displaying some properties of a DNA solution against an external variable producing DNA unwinding. The most common external variable is the temperature, although the process can also be observed at extremes of pH, which decreases in the dielectric constant of the aqueous medium, when exposed to amides, urea, and similar solvents. The DNA property of optical absorbance can be monitored at approximately 260 nm, while increasing the temperature and normalizing the absorbance change in an appropriate way. Thus, a plot of the fraction of broken base-pairs (bps) versus temperature is obtained. In the double helix, disruption of the ordered state, with its stacked-base pairs, leads to less frequent contact between the bases and an increased absorbance [7].

Since the pioneering work of Zimm [8], and the appearance of the nearest-neighbor (NN) model, which assumes that the stability of a given bp depends on the identity and orientation of the neighboring base pair, several theoretical and experimental papers on DNA thermodynamics have appeared. The experimental works provide the complete thermodynamics library of all 10 Watson-Crick DNA nearest-neighbor interactions. Good descriptions of a sampling of experimental techniques used for this purpose, and the principal thermodynamics libraries available in the literature, are described in [9]. Even though the thermodynamic sets given in these articles disagree on a number of issues, they show how these thermodynamic data can be used to calculate the stability of the structure from the knowledge of its base sequence.

The particular differences are between DNA polymers and oligonucleotides, and in the salt dependence of nucleic acid denaturation. As a result of these works, it is well established that



is more stable than



Besides, theoretical efforts have been made to calculate stacking energies *ab initio*; for a review consult reference [10].

Recent research has emphasized the role of the large amplitude fluctuations that precede the transition and the intrinsically nonlinear mechanisms which are needed to describe such fluctuations. This description was introduced because, experimentally a purified DNA sample containing a unique sequence and length is found to exhibit distinct multi step melting [7].

The early theoretical treatment by Poland and Scheraga (PS) is inspired in the ferromagnetic 1D Ising model [11]. To describe the melting curve, *i.e.*, the fraction of bonded or unbonded base-pairs as a function of the temperature in the solution, the PS theory consists in obtaining the partition function of a DNA chain formed by pairs, characterized by one of the two states, bounded or unbonded, related with the original model as spins up or down. In the PS method in order to take into account the cooperative nature of the helix-coil transition, is introduced an adjustable parameters, called the nucleation parameter with values between 0 and 1. This model reproduces a crossover between the two different regimes, but no thermo-dynamical transition. The relative tendencies of the system to occupy one of the two states, were introduced explicitly in terms of free enthalpies, and their temperature dependence. Although the choice of such enthalpies is difficult, the method has proved being useful in describing some aspects of DNA denaturation [11]. Recently, in order to take into account finer aspects in the denaturation, as the entropic and torsional effects in DNA chain during the transition, some authors have proposed additional adjustable parameters to solve the description for some heterogeneous sequences [7]. Understanding of this remarkable one-dimensional cooperative phenomenon in terms of a Hamiltonian model with independent parameters, remained an outstanding problem. Moreover for some DNAs the PS description fails [22].

In the present paper, we propose a new theoretical model to describe the denaturation process of DNA as a Markov process, taking into account the probability of finding each nearest stacked neighbor. This model extends the stochastic matrix method (SMM) [12–14], used to describe the glass transition in strong glasses, to the study of configurational changes in the denaturation process and to the prediction of T_m of DNA for polynucleotides [15].

2. Model and Definitions

Our theoretical model has important differences from the treatment by PS. First, the statistical treatment is based on a first-order Markov process, an easier mathematical framework than the PS algorithm. Finally, our model is based on the fact that the thermodynamic values for the ten possible dimers are known, and as a consequence, the external conditions as pH, salt concentration solution, the role of the large amplitude fluctuations and the intrinsically nonlinear mechanisms are included.

In the stochastic matrix method, the process of observing the configuration of bps in DNA can be described by a matrix (M) acting on an initial vector v_0 (which characterizes the initial condition of the system), if the matrix components are the probability of having a bp neighboring another one in a specific configuration, and if the vector components represent the probabilities of having a bp in the configuration. The probability of having some configuration of bps is modeled by n successive applications of the matrix M on an initial vector v_0 . After n applications, the final configuration of the system can be written as a linear combination of the eigenvectors associated with M , *i.e.*, $v_n = \sum_{m=1}^n a_m \lambda_M^n e_m$, where e_m is the eigenvector M with eigenvalue λ_M^n , and a_m is the projection of v_0 onto e_m .

A matrix with all the columns normalized to one, as M , has the property that at least one eigenvalue is one, while the real part of all the rest is less than one. This result allows us to assert that only the eigenvectors with eigenvalues equal to one, survive after successive applications of M onto v_0 . If we assume that M has one eigenvector e_1 , with eigenvalue equal to one, then in the limit of a large n , v_∞ converges to e_1 , with $a_1 = 1$, due to a conservation of probability. Therefore, this means that the configuration attains a steady statistical regime represented by e_1 . The explicit form of this eigenvector is obtained by solving the system of equations:

$$(M - 1)e_1 = 0, \quad (1)$$

which enables us to calculate the probability of any configuration in the system.

To construct the stochastic matrix describing the melting behavior of DNA, we first need to define the units. These units must be given by four combinations: A and T, T and A, G and C, and C and G. The bps can be bonded or unbonded. This can be represented as $\uparrow A \cdot T \downarrow$, $\uparrow T \cdot A \downarrow$, $\uparrow G \cdot C \downarrow$, $\uparrow C \cdot G \downarrow$, and $\uparrow A \ T \downarrow$, $\uparrow T \ A \downarrow$, $\uparrow G \ C \downarrow$, $\uparrow C \ G \downarrow$, where the dot represents the existence of the hydrogen bonds between the bps, and the absence of the dot represents the unbonded bps [16]. These eight units give 64 different combinations of base-pair stacking, where each site of the matrix represents the probability of finding a specific configuration of each duplex.

The 64 different combinations can be displayed as an 8×8 matrix, namely:

$$\left(\begin{array}{cccccccc} \uparrow G \cdot C \downarrow & \uparrow C \cdot G \downarrow & \uparrow A \cdot T \downarrow & \uparrow T \cdot A \downarrow & \uparrow G C \downarrow & \uparrow C G \downarrow & \uparrow A T \downarrow & \uparrow T A \downarrow \\ \uparrow G \cdot C \downarrow & \uparrow G \cdot C \downarrow \\ \uparrow C \cdot G \downarrow & \uparrow C \cdot G \downarrow \\ \uparrow G \cdot C \downarrow & \uparrow C \cdot G \downarrow & \uparrow A \cdot T \downarrow & \uparrow T \cdot A \downarrow & \uparrow G C \downarrow & \uparrow C G \downarrow & \uparrow A T \downarrow & \uparrow T A \downarrow \\ \uparrow A \cdot T \downarrow & \uparrow A \cdot T \downarrow \\ \uparrow G \cdot C \downarrow & \uparrow C \cdot G \downarrow & \uparrow A \cdot T \downarrow & \uparrow T \cdot A \downarrow & \uparrow G C \downarrow & \uparrow C G \downarrow & \uparrow A T \downarrow & \uparrow T A \downarrow \\ \uparrow T \cdot A \downarrow & \uparrow T \cdot A \downarrow \\ \uparrow G \cdot C \downarrow & \uparrow C \cdot G \downarrow & \uparrow A \cdot T \downarrow & \uparrow T \cdot A \downarrow & \uparrow G C \downarrow & \uparrow C G \downarrow & \uparrow A T \downarrow & \uparrow T A \downarrow \\ \uparrow G C \downarrow & \uparrow G C \downarrow \\ \uparrow G \cdot C \downarrow & \uparrow C \cdot G \downarrow & \uparrow A \cdot T \downarrow & \uparrow T \cdot A \downarrow & \uparrow G C \downarrow & \uparrow C G \downarrow & \uparrow A T \downarrow & \uparrow T A \downarrow \\ \uparrow C G \downarrow & \uparrow C G \downarrow \\ \uparrow G \cdot C \downarrow & \uparrow C \cdot G \downarrow & \uparrow A \cdot T \downarrow & \uparrow T \cdot A \downarrow & \uparrow G C \downarrow & \uparrow C G \downarrow & \uparrow A T \downarrow & \uparrow T A \downarrow \\ \uparrow A T \downarrow & \uparrow A T \downarrow \\ \uparrow G \cdot C \downarrow & \uparrow C \cdot G \downarrow & \uparrow A \cdot T \downarrow & \uparrow T \cdot A \downarrow & \uparrow G C \downarrow & \uparrow C G \downarrow & \uparrow A T \downarrow & \uparrow T A \downarrow \\ \uparrow T A \downarrow & \uparrow T A \downarrow \end{array} \right) \tag{2}$$

where

$$\begin{array}{c} \uparrow G \cdot C \downarrow \\ \uparrow G \cdot C \downarrow \end{array}$$

represents the probability of having a bonding $\uparrow G \cdot C \downarrow$ neighboring a bonding

$$\begin{array}{c} \uparrow G \cdot C \downarrow, \uparrow G C \downarrow \\ \uparrow G C \downarrow \end{array}$$

represents the probability of having a bonding $\uparrow G \cdot C \downarrow$ neighboring an unbonded $\uparrow G C \downarrow$, etc. These stacking processes are in three dimensions, and this information must be included in the stacking energy, as should the information concerning the properties of the solvent in which the transition is carried out.

Based on the NN model, each configuration is proportional to two factors: the concentration of the bps, and its Boltzmann factor. The first one depends on the sequence, and the second one, on the possible configurations. This last factor involves the Gibbs free energy of each configuration, namely, $\exp[\Delta G_{MN}/k_B T]$ (where M and N stand for A, T, G, C, and MN represents the stacked bps in a single strand in the direction 5'–3'). If each configuration is proportional to its stability constant, it means that if the probability of some bp is found bond or unbonded at a given temperature, this bp is unable to change its configuration at this temperature.

The eigenvector with eigenvalue equal to one of matrix (2) is a vector with eight components that gives us the probability of finding, at a fixed temperature, the following configurations in the system: $\uparrow G \cdot C \downarrow$, $\uparrow C \cdot A \downarrow$, $\uparrow A \cdot T \downarrow$, $\uparrow T \cdot A \downarrow$, $\uparrow G C \downarrow$, $\uparrow C G \downarrow$, $\uparrow A T \downarrow$, and $\uparrow T A \downarrow$. The sum

$$\uparrow G C \downarrow + \uparrow C G \downarrow + \uparrow A T \downarrow + \uparrow T A \downarrow$$

gives us the probability of denatured bps in the system. Setting this sum equal to 1/2, T_m is obtained for DNA.

In the next section, we shall give three examples that enable us to show how to use the matrix (2) to obtain the T_m

behavior of the melting transition, and how to obtain the for DNA. These examples are focused on the study of polynucleotides.

3. Results and Comparison with Experiment

In the first two subsections, we shall discuss the application of the method and the results for periodic chains. Simple expressions, for melting curves and melting temperatures of DNA, composed for periodic distributions of GC or AT pairs, can be derived. It is useful to discuss such cases in detail because a number of general qualitative similarities and differences, which result from different distributions, already show up in these cases. One can use these results to check the validity of the solutions obtained for more complicated distributions. As we shall show, the more complicated cases can always be reduced to a simple one, under specific conditions in sequence concentrations. In the last subsection, we discuss the case in which a random sequence of bps is imposed.

3.1. Poly(dM)·poly(dN)

In our first example, we study the denaturation transition for poly(dM)·poly(dN). For simplicity, we will work here with poly(dG)·poly(dC). The results obtained are also applicable to poly(dA)·poly(dT) using the correct energies for the duplex involved in the Boltzmann factor. To obtain the behavior of the transition, we first construct the stochastic matrix. Then, we solve equation (1) for this particular case. To calculate the T_m , we proceed as follows. Obtaining the eigenvector with eigenvalue equal to one, we can find the probability of natured and denatured bps. Using the definition of T_m , we set the probability of denatured bps to 1/2, and finally we solve the equation for T_m to find the melting temperature.

For this particular case only the terms



remain in matrix (2), and the fraction of this bps in the system is equal to 1. Taking this considerations into account, matrix (2) is reduced to 2×2 matrix,

$$\begin{pmatrix} \begin{matrix} \uparrow & \text{G} \cdot \text{C} & \downarrow & \uparrow & \text{G} & \text{C} & \downarrow \\ & \text{G} \cdot \text{C} & & & \text{G} \cdot \text{C} & & \\ \uparrow & \text{G} & \text{C} & \downarrow & \uparrow & \text{G} & \text{C} \\ & \text{G} \cdot \text{C} & & & \text{G} & \text{C} & \downarrow \end{matrix} \end{pmatrix}. \quad (3)$$

Inserting the energetic contributions in matrix 3 we find,

$$\begin{pmatrix} \xi & \xi\mu \\ \xi\mu & 1 \end{pmatrix}, \quad (4)$$

where $\xi \equiv \exp[-\Delta G_{GG}/k_B T]$, the free energy for the dimer



and $\mu \equiv \exp[\Delta G_h/k_B T]$, which involves the free energy for the hydrogen bond between G and C. If the hydrogen bond of a bp is broken in a duplex, we have $\Delta G_{GG} - \Delta m G_h$, and the Boltzmann factor for this configuration is given by $\xi\mu$. If two hydrogen bonds of the bps involved are broken, its Gibbs free energy is zero.

After normalizing each column of matrix (4) we have:

$$\begin{pmatrix} \frac{1}{1+\mu} & \frac{\xi\mu}{1+\xi\mu} \\ \frac{\mu}{1+\mu} & \frac{1}{1+\xi\mu} \end{pmatrix}. \quad (5)$$

The explicit form of the eigenvector with eigenvalue one is obtained by solving Eq. (1) with M given by matrix 2. This process yields the following vector:

$$\begin{pmatrix} \uparrow \text{G} \cdot \text{C} \downarrow \\ \uparrow \text{G} \text{C} \downarrow \end{pmatrix} = \begin{pmatrix} \frac{\xi(1+\mu)}{1+\xi+2\mu} \\ \frac{1+\mu\xi}{1+\xi+2\mu} \end{pmatrix}. \quad (6)$$

Vector (6) gives us the probability of finding $\uparrow \text{G} \cdot \text{C} \downarrow$, and $\uparrow \text{G} \text{C} \downarrow$ base-pairs in poly(dG)·poly(dC) at any temperature. This is the well known solution to a Markov chain of two states. Using vector (6), we can obtain the behavior of the melting transition expressed in terms of the fraction of open bps θ , given by the second component of vector (6) divided by the sum of the first and second components. Because the conservation of probability, the sum of both terms is equal to 1, θ is given just by

$$\theta = \frac{1 + \xi\mu}{1 + \xi + 2\xi\mu}. \quad (7)$$

To obtain configurational changes predicted by Eq. (7), we have to insert the experimental values of the stacking free energies and the free energy of the hydrogen bonds. Because we assume that each duplex is in a state of thermodynamic equilibrium, ΔG_{GG} , and ΔG_h can be calculated from

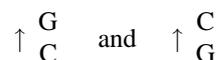
the empirical relations $\Delta G_{GG} = \Delta H_{GG} - T\Delta S_{GG}$, and $\Delta G_h = \Delta H_h - T\Delta S_h$; where ΔH_{GG} and ΔS_{GG} stand for the enthalpy and entropy due to stacking, and ΔG_h and ΔS_h , for the enthalpy and entropy of the hydrogen bond between G and C.

In Fig. 1, for poly(dG)·poly(dC), θ is plotted vs. temperature. The experimental values used in Eq. 7 are $\Delta H_{GG} = -8993 \text{ kcal/mol}$, $\Delta S_{GG} = -24.85 \text{ e.u.}$ [20], $\Delta H_h = -5.8 \text{ kcal/mol}$ and $\Delta S_h = -16 \text{ e.u.}$ [17]. The experimental values for the stacking enthalpies and entropies throughout this article are taken from Table II in Ref. 20 (For an excellent description and discussion about the thermodynamics libraries see Ref. 9). The experimental values for hydrogen enthalpy and entropy in all our calculations are also taken from Ref. 17. In Fig. 1, it is observed that θ has the same shape as the experimental curves as a function of the temperature.

Now, if we want to calculate T_m , we only have to set θ to one half. This condition is reached when $\xi = 1$, that means $\Delta G_{GG} = 0$. Then, $T_m = \Delta H_{GG}/\Delta S_{GG}$, is the expected result.

3.2. Poly[(dM-dN)·(dM-dN)]

Now, as in the preceding example, by using matrix (2), we describe the melting behavior and obtain T_m for poly[(dG-dC)·(dG-dC)]; the results are also applicable to poly[(dA-dT)·(dA-dT)], poly[(dA-dT)·(dG-dC)], etc. For this particular case, in matrix (2) only, the terms



remain, and the fraction of the pairs involved are equal to 1/2. Matrix (2) is reduced to a 4×4 matrix, namely:

$$\begin{pmatrix} \begin{matrix} \uparrow & \text{G} \cdot \text{C} & \downarrow & \uparrow & \text{C} \cdot \text{G} & \downarrow & \uparrow & \text{G} & \text{C} & \downarrow & \uparrow & \text{C} & \text{G} \\ & \text{G} \cdot \text{C} & & & \text{G} \cdot \text{C} & & & \text{G} & \text{C} & & & \text{C} & \text{G} \\ \uparrow & \text{G} \cdot \text{C} & \downarrow & \uparrow & \text{C} \cdot \text{G} & \downarrow & \uparrow & \text{G} & \text{C} & \downarrow & \uparrow & \text{C} & \text{G} \\ & \text{G} \cdot \text{C} & & & \text{C} \cdot \text{G} & & & \text{G} & \text{C} & & & \text{C} & \text{G} \\ \uparrow & \text{G} & \text{C} & \downarrow & \uparrow & \text{G} & \text{C} & \downarrow & \uparrow & \text{G} & \text{C} & \downarrow & \uparrow & \text{G} & \text{C} \\ & \text{G} \cdot \text{C} & & & \text{C} \cdot \text{G} & & & \text{G} & \text{C} & & & \text{C} & \text{G} \\ \uparrow & \text{G} \cdot \text{C} & \downarrow & \uparrow & \text{C} \cdot \text{G} & \downarrow & \uparrow & \text{G} & \text{C} & \downarrow & \uparrow & \text{C} & \text{G} \\ & \text{G} \cdot \text{C} & & & \text{C} \cdot \text{G} & & & \text{G} & \text{C} & & & \text{C} & \text{G} \\ \uparrow & \text{C} & \text{G} & \downarrow & \uparrow & \text{C} & \text{G} & \downarrow & \uparrow & \text{C} & \text{G} & \downarrow & \uparrow & \text{C} & \text{G} \\ & \text{C} & \text{G} & & & \text{C} & \text{G} & & & \text{C} & \text{G} & & & \text{C} & \text{G} \end{matrix} \end{pmatrix}. \quad (8)$$

Introducing the fraction of appearance for each configuration and their Boltzmann factor, we have

$$\frac{1}{2} \begin{pmatrix} 0 & \varepsilon & 0 & \varepsilon\mu \\ \eta & 0 & \eta\mu & 0 \\ 0 & \varepsilon\mu & 0 & 1 \\ \eta\mu & 0 & 1 & 0 \end{pmatrix}, \quad (9)$$

where $\eta \equiv \exp[-\Delta G_{GC}/k_B T]$, $\varepsilon \equiv \exp[-\Delta G_{CG}/k_B T]$, and μ has the same meaning as in the preceding section.

After normalizing each column of matrix (9) and solving Eq. (2) with M given by Eq. (9), a vector of four components is found. The first and second components of this vector give us the probability of finding $\uparrow \text{G} \cdot \text{G} \downarrow$ and $\uparrow \text{C} \cdot \text{G} \downarrow$, the

probability of the closed duplex. The sum of the last two, $\uparrow G C \downarrow + \uparrow C G \downarrow$, give us the probability of finding the totally open bps, namely,

$$\theta = \frac{4 + \mu(2 + \eta(3 + \mu)) + \varepsilon(3 + \mu + 2\eta\mu)}{2(2 + \varepsilon + \eta + \mu(1 + 3\eta + \varepsilon(3 + \eta))) + \mu^2(\varepsilon + \eta + 2\varepsilon\eta)} \quad (10)$$

In Fig. 1, Eq. (10) is plotted. The shape obtained is that observed experimentally, in fact, it is very close to that obtained in the preceding section for poly(dG)·poly(dC). More details about the experimental values used in Eq. (10) are given in the figure caption.

To calculate T_m , it is necessary to impose $\theta = 1/2$, Eq. (10) is equal to 1/2 when,

$$2 - \varepsilon - \eta + \mu - \varepsilon\eta\mu = 0. \quad (11)$$

Equation (11) depends on the hydrogen bond parameters as well as the stacking ones. In this case, the role of the stacking energies is as fundamental as that of the hydrogen bonding energies, and the competition between their ΔS_s and their ΔH_s governs the melting behavior.

The theoretical value for T_m obtained using Eq. (11), and the experimental data of Delcourt [20], and Newmark [17], is 115°C. The experimental melting temperature for this sequence is 111.83°C [18], while Delcourt obtained 107.99° [20]. (The Delcourt values were obtained by $T_m = \sum f_{MN} \Delta G_{MN}$). Our result is closer to the experimental one. Presumably, our theoretical value could be better if the thermodynamical parameters used for the hydrogen bond were obtained for the same solvent as the stacking ones.

3.3. Poly[d(M,N)-d(M,N)]

To solve the problem of poly[d(M,N)·d(M,N)], as in the preceding subsection, for simplicity we will work with poly[(dG,dC)·(dG,dC)], and all the results obtained are applicable to poly[(dA,dT)·(dA,dT)], as well as poly[(dG,dC)·(dA,dT)], etc. In this particular case, matrix (2) is reduced to a 4×4 matrix, Eq. (8). Introducing the fraction of each duplex and their Boltzmann factor the 4×4 matrix is given by,

$$\begin{pmatrix} f_{GG} \xi & f_{CG} \varepsilon & f_{GG} \xi \mu & f_{CG} \varepsilon \mu \\ f_{GC} \eta & f_{GG} \xi & f_{GC} \eta \mu & f_{GG} \xi \mu \\ f_{GG} \xi \mu & f_{CG} \varepsilon \mu & f_{GG} & f_{CG} \\ f_{GC} \eta \mu & f_{GG} \xi \mu & f_{GC} & f_{GG} \end{pmatrix}, \quad (12)$$

where ξ , ε , η , and μ have the same meaning as in the preceding sections. Following the same procedure as in the two preceding sections, after normalizing Eq. (12), one can obtain θ . Even though the steps to obtain θ for this particular case are straight forward, the explicit equation contains a large number of factors. For this reason the full expression is shown in Appendix, in Eq. (A.1).

In Fig. 1, θ , given by Eq. (A.1), is plotted for $f_{GG} = 0.3$, $f_{CG} = 0.3$, and $f_{GC} = 0.4$. The details of the experimental values, used in this plot, are given in the figure caption.

The shape of the curve predicted by the theory is the expected. It is important to remark that with any combination of f_{CG} and f_{GC} the denatured curves predicted by Eq. (A.1), are in between those obtained for poly(dG)·poly(dC), and poly[(dG-dC)·(dG-dC)].

In Fig. 2, T_m is plotted vs. the fraction of duplex. This fraction refers to the fraction in a single strand, and the difference sequence in a vertical stacking. This plot takes

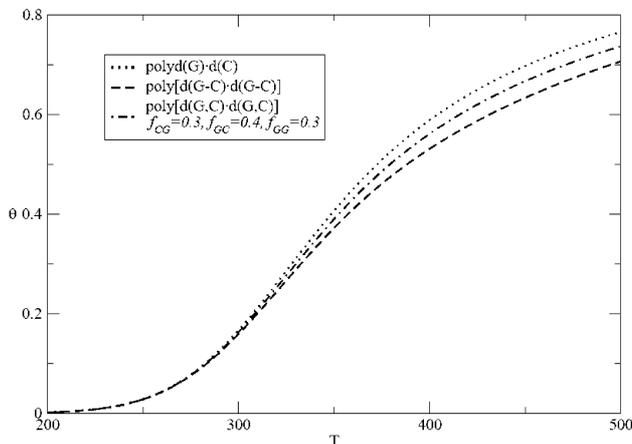


FIGURE 1. Plot of θ for three different cases, poly(dG)·poly(dC) given by Eq. (5), [(dG-dC)·(dG-dC)] given by Eq. (10), and poly[(dG,dC)·(dG,dC)] given by Eq. (A.1), with $f_{GG} = 0.3$, $f_{CG} = 0.3$, and $f_{GC} = 0.4$. The two first cases are particular cases of the last one when $f_{GC} = f_{CG} = 0$, and when $f_{GC} = 0$, respectively. The experimental thermodynamic hydrogen bond values were taken from Ref. 17. The experimental thermodynamic nearest-neighbor values were taken from Ref. 7.

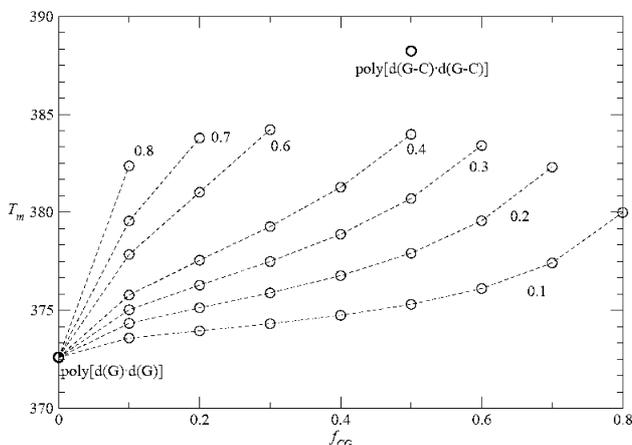
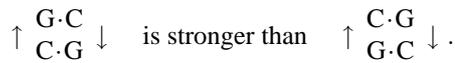


FIGURE 2. Plot for poly(dG,dC)·poly(dG,dC) of T_m vs. all possible combinations of f_{GG} , f_{CG} , and f_{GC} . Because $f_{GG} + f_{GC} + f_{CG} = 1$, we only have two independent parameters; to plot T_m in two dimension, we set constant f_{GC} , whose value is indicated in each case. Even though [G+C] remains constant in all the cases, the plot shows the strong dependence of the sequences. The two extreme values are given by poly(dG)·poly(dC) ($f_{GG} = 1$), the least stable of all the conformations, and by poly[(dG-dC)·(dG-dC)] ($f_{CG} = f_{GC} = 1/2$, and $f_{CG} = 0$), the most stable sequence. The thermodynamics experimental values were taken from Refs. 7 and 17.

into account not only the total fraction of G and C, [G+C], but also the variation of each duplex in the sequence. Even though the total fraction of G and C remains constant for each point plotted in Fig. 2, the melting temperature is sequence dependent. We predict with our model, an important variation for the single point obtained by the linear Marmur-Doty relation [21], and the entire range of possible melting temperatures around 16°C. The extreme values are given by poly(dG)·poly(dC), and by poly[(dG-dC)·(dG-dC)].

It is important to remark that one of the interesting features observed in Fig. 2, is the case when the values of f_{GC} and f_{CG} are switched: and T_m changes. For example, the T_m predicted for a DNA sequence with the fraction $f_{GC} = 0.1$, $f_{CG} = 0.8$ ($f_{GG} = 0.1$) is 106.84 °C, while for the fraction $f_{GC} = 0.8$, $f_{CG} = 0.1$ ($f_{GG} = 0.1$) 109.21 °C is obtained. This feature is observed because



Finally, setting in Eq. (A.1) $f_{CG} = f_{GC}$ and $f_{CG} = 1$, this expression reduces to Eq. (7). Setting $f_{CG} = f_{GC} = 1/2$ and $f_{CG} = 0$, Eq. (10) is obtained.

4. Conclusions

To summarize, we used the stochastic matrix method to study the melting of DNA, and its melting temperature dependence on the fractional composition in a single strand. The elements of the stochastic matrix are the probabilities that any base-pair has of a bonded or an unbonded neighbor. If one assumes that the stochastic matrix has an eigenvector with eigenvalue equal to one, the possible configurations of the system are fixed by this eigenvector. In fact, this vector is the probability of finding any bonded or unbonded base. Once we have this probability, we are able to obtain an analytic expression for the fraction of broken base-pairs as a function of temperature. Besides, setting to 1/2 the expression for this probability, gives us an analytical expression to calculate the melting temperature. This theoretical method is free of adjustable parameters, and to carry out the comparison with experimental data, only the DNA nearest-neighbor thermodynamics set, as well as the hydrogen bond parameter are needed. The shapes predicted by our theoretical expression for temperature dependence of the denatured base-pairs are the expected ones for a DNA chain. The T_m obtained theoretically are in good agreement with the experimental values. The T_m extreme values are obtained when $f_{CG} = 1$ and when $f_{CG} = 0$, and $f_{CG} = f_{GC} = 1/2$. The first is for poly(dG)·poly(dC) and is the lowest value. The second extreme gives the strongest possible chain, poly[(dG-dC)·(dG-dC)]. The rest of the possible

combinations of fractions are in between. The maximum difference in T_m for chains with the same fraction of [G+C] is around 16°C. Thus, instead of having a single point as in the Marmur-Doty relation, for this particular case we obtain a set of points with T_m values between the extremes. Our theory can be used to predict any combination [G+C] and [A+T].

Finally, it is important to say that this method can be used to describe the unfolding behavior for peptides and proteins.

Appendix: poly[d(G,C)·d(G,C)]'s θ

In this appendix we shall show an expression for poly[d(G,C)·d(G,C)]'s θ . For this particular case,

$$\theta = N/D. \tag{A.1}$$

where

$$\begin{aligned} N = & (a^2 - bc + S_1S_2 - a(S_1 + S_2)) \\ & \times (S_3(f_{CG} + S_4) - f_{GG}S_4) \\ & + (c(bc + a(S_1 + S_2 - a))S_3) \\ & + \mu^2(a_3 - abc - bcS_1 - a_2S_2)S_4), \end{aligned} \tag{13}$$

$$\begin{aligned} D = & (a^2 - bc + S_1S_2 - a(S_1 + S_2)) \\ & \times (S_3(f_{CG} + S_4) - f_{GG}S_4) - cS_1S_2(f_{GG} - S_3) \\ & - aS_1S_2(f_{CG} - f_{GG} + S_3) \\ & - b(c(f_{CG}S_1 - f_{GG}S_2 + S_2S_3) + f_{CG}S_1S_2) \\ & + a_2\mu(f_{CG}S_1 + S_2(S_3 - f_{GG})) \\ & + (c(bc + a(S_1 + S_2 - a))S_3) \\ & + S_4\mu^2(a_3 - abc - bc(1 - S_1) - a_2S_2) \\ & - \mu^3(a_2 - bc)(aS_2 - cS_1) \end{aligned} \tag{14}$$

$$S_1 = (1 + \mu)(a + b)$$

$$S_2 = (1 + \mu)(a + c)$$

$$S_3 = \mu(a + b) + f_{GG} + f_{GC}$$

$$S_4 = \mu(a + c) + f_{GG} + f_{CG} \tag{15}$$

and

$$a = f_{GG}\xi,$$

$$b = f_{GC}\eta,$$

$$c = f_{CG}\varepsilon.$$

1. R.K. Saiki *et al.*, *Science* **239** (1988) 487.
2. S.P.A. Fodor *et al.*, *Nature* **364** (1993) 555.
3. S.M. Freier, *Antisense Research and Applications* S.T. Crooke and B. Lebleu Eds. (CRC Press, Boca Raton, FL, 1993) p. 67.
4. E.M. Southern, *J. Mol. Biol.* **98** (1975) 503.
5. G. Steger, *Nucleic Acids Res.* **22** (1994) 2760.
6. V.A. Bloomfield, *Physical chemistry of nucleic acids* (Harper and Row, New York, 1974).
7. M. Wartell and A.S. Benight, *Phys. Rep.* **126** (1985) 67.
8. D.M. Crothers and B.H. Zimm, *J. Mol. Biol.* **9** (1964) 1.
9. Jr. SantaLucia, *Proc. Natl. Acad. Sci. USA.* **95** (1998) 1460.
10. P. Hobza and J. Šponer, *Chem. Rev.* **99** (1999) 3247.
11. D. Poland and H.A. Scheraga, *Theory of Helix-Coil Transition in Biopolymer* (Academic Press, New York, 1970).
12. R. Kerner, *Phys. B* **215** (1996) 267.
13. R.A. Barrio, R. Kerner, M. Micoulaut, and G.G. Naumis, *J. Phys.: Condens Matter* **9** (1997) 9219.
14. R. Kerner and G.G. Naumis, *J. Phys.: Condens Matter* **12** (2000) 1641.
15. L. Dagdug and E. Vázquez-Contreras, *Rev. Mex. Fis.* **48** S1 (2002) 168.
16. Arrows designate the direction of the sugar-phosphate chain, from the C₃' atom of a deoxyribose unit to the C₅' atom of the next deoxyribose adjacent to and on either side of the phosphodiester linkage. Sometimes nearest-neighbor base pairs are represented with a slash separating strands in an antiparallel orientation (e. g., AC/TG means 5'-AC-3' Watson-Crick bases paired with 3'-TG-5' or $\uparrow \begin{matrix} A \cdot T \\ C \cdot G \end{matrix} \downarrow$ in the notation used throughout this paper).
17. R.A. Newmark and C.R. Cantor, *J. Amer. Chem. Soc.* **90** (1968) 5010.
18. R.D. Wells, J.E. Larson, R.C. Grant., B.E. Shortle, and C.R. Cantor, *J. Mol. Biol.* **54** (1970) 465.
19. K.J. Breslauer, R. Frank, H. Blocker, and L.A. Marky, *Proc. Natl. Acad. Sci. U.S.A.* **83** (1986) 3746.
20. S.G. Delcourt and R.D. Blacke, *J. Biol. Chem.* **266** (1991) 15160.
21. J. Marmur and P. Doty, *J. Mol. Biol.* **3** (1962) 109.
22. H. Tachibana, S. Ueno-Nishio, O. Gotoh, and A. Wada, *J. Biochem.* **92** (1982) 623.