Theoretical predictions of the melting temperature for DNA using the stochastic matrix method

L. Dagdug
Departamento de Física, Universidad
Autónoma Metropolitana-Iztapalapa
Apdo. Post. 55-534, 09340 México, D.F., México
e-mail: dll@xanum.uam.mx

E. Vázquez-Contreras
Instituto de Química, Departamento de Bioquímica, Universidad Nacional Autónoma de México
04510 México, D.F., México

Recibido el 19 de marzo de 2001; aceptado el 21 de noviembre 2001

In this paper we extend the ideas used to describe the glass transition in strong glasses using the stochastic matrix method to obtain the melting temperature for DNA. To this end, we take the purines and pyrimidines as principal entities, the simplest ones. The main result is the prediction of the melting temperature of poly(dpurine)-poly(dpyrimidine) and poly(dpurine-dpyrimidine)-poly(dpurine-dpyrimidine) without adjustable parameters and with excellent agreement with experimental data.

Keywords: DNA; melting temperature; helix-coil transition; nearest-neighbor thermodynamics; stochastic matrix method

En este artículo se extienden las ideas utilizadas para describir la transición vitrea de vidrios fuertes utilizando el método de la matriz estocástica, para obtener la temperatura de desnaturalización del ADN. Para llevar a cabo nuestro propósito utilizamos purinas y pirimidinas como nuestra unidad fundamental, la más simple posible. El resultado principal que se obtiene de este análisis, es la predicción de las temperaturas de desnaturalización del poly(dpurina)-poly(dpirimidina) y del poly(dpurina-dpirimidina)-poly(dpurina-dpirimidina), sin utilizar parámetros ajustables y con un excelente ajuste a los valores obtenidos experimentalmente.

Descriptores: ADN; temperatura de desnaturalización; transición; termodinámica de primeros vecinos; método de la matriz estocástica

PACS: 87.10+e

Dedicated to Leopoldo García-Colín Scherer who taught me how to confront scientific problems as a scientist. (L. Dagdug)

1. Introduction

Nucleic acids consist of a chemical sequence of their fundamental components, nucleotides, analogous to the sequence of linked amino acids constituting a protein. Nucleotides comprise three elements: a heterocyclic ring containing nitrogen, a five-carbon sugar in ring form, and a phosphate group. The bases fall into two classes, pyrimidines (Py) and purines (Pu); the former have a six-membered ring and the latter consist of two fused six- and five-membered rings. Each nucleic acid commonly contains four different types of bases, two of each ring form, although the two types of nucleic acid, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), differ in one of the pyrimidines presented. Both DNA and RNA contain the same two purines, adenine and guanine. DNA has the pyrimidines cytosine and thymine; in RNA the thymine is replaced by uracil. In the primary structure of DNA, each nucleoside is joined by a phosphodiester from its -hydroxyl group to the -hydroxyl group of one neighbor and by a second phosphodiester from its -hydroxyl group to the -hydroxyl of its other neighbor [1].

In general, DNA consists of two very long polynucleotide chains wound around each other to form a double helix. Watson and Crick [2] suggested that this could be accounted for if a purine from one polynucleotide chain always partners with a pyrimidine from the other chain; thus, both purine-purine and pyrimidine-pyrimidine interactions would be prohibited. They suggested that the two strands would be held together by hydrogen bonding, and would not be covalently linked. The building of molecular models explained this by suggesting that hydrogen bonding could occur between the nitrogenous bases such that the purine guanine could bond only to the pyrimidine cytosine, and not to the thymine. Similarly, adenine and thymine would form hydrogen bonds only with each other. Thus a thymine of one chain would always oppose an adenine. The adenine-thymine (A·T) base pair (bp) has two hydrogen bonds, and the guanine-cytosine (G·C) pair has three. The two bases comprising each pair are said to be complementary. This model requires that the two-polynucleotide chains run in opposite directions, antiparallel. Since pairing occurs in this way, the amount of adenine and thymine should be equal, and quantities of cytosine and guanine should likewise be equal to each other. Also, the ratio between purines and pyrimidines is one. This last result was an important piece of information for the development of the structure of DNA and came from the work of Chargaff and co-workers [3].

The noncovalent forces that stabilize the double helix are disrupted by heating or by exposure to a low salt concentration (osmotic shock). The two strands of the double helix separate entirely when all the hydrogen bonds between them are broken. The process of strand separation is called
denaturation or melting. The mid point of the temperature range over which the strands of DNA separate is called the melting temperature, denoted $T_m$. The investigation of this phenomenon has provided information and insight on the interactions governing DNA unwinding, the influence of base pair sequence on regional unwinding, and the influence of the solvent on DNA stability [13–16].

A DNA melting curve is generally a two-dimensional plot displaying some property of a DNA solution against an external variable producing DNA unwinding. The external variable is commonly the temperature, but melting can also be brought about by extremes of pH, decreases in the dielectric constant of the aqueous medium by alcohol, ketones, etc., and exposure to amides, ureas and similar solvents [4].

Studies of the structure and dynamics of DNA are vital to understand the mechanism of how the genetic code is expressed, the processes of DNA replication and transcription, DNA-protein recognition, DNA-drug interactions and more [5].

Accurate prediction of DNA thermal denaturation is important for several bimolecular techniques including PCR and exposure to amides, ureas and similar solvents [4]. The investigation of this phenomenon has provided information and insight on the interactions governing DNA unwinding, the influence of base pair sequence on regional unwinding, and the influence of the solvent on DNA stability [13–16].

A DNA melting curve is generally a two-dimensional plot displaying some property of a DNA solution against an external variable producing DNA unwinding. The external variable is commonly the temperature, but melting can also be brought about by extremes of pH, decreases in the dielectric constant of the aqueous medium by alcohol, ketones, etc., and exposure to amides, ureas and similar solvents [4].

Studies of the structure and dynamics of DNA are vital to understand the mechanism of how the genetic code is expressed, the processes of DNA replication and transcription, DNA-protein recognition, DNA-drug interactions and more [5].

Accurate prediction of DNA thermal denaturation is important for several bimolecular techniques including PCR [6], sequencing by hybridization [7], antigen targeting [8], and Southern blotting [9]. In these techniques, choice of a non-optimal sequence or temperature can lead to amplification or detection of wrong sequences [10]. Furthermore, the knowledge of the sequence dependence of DNA melting is important for the understanding of the details of DNA replication, mutation, repairing, and transcription [11, 12].

In this paper we extend the ideas to describe the glass transition in strong glasses using the stochastic matrix method (SMM) to predict the melting temperatures of DNA [17–20]. In Sec. 2, we present the stochastic matrix method. In Sec. 3, we use the theoretical ideas presented in Sec. 2 to describe the melting process in DNA; for example, we obtain the $T_m$ of poly(dpurine)-poly(dpyrimidine) and poly(dpurine-dpyrimidine)-poly(dpurine-dpyrimidine). Finally, in Sec. 4, we make some remarks on the nature of these results.

2. The stochastic matrix method

Observation of the configuration of a system can be described by a matrix (M) acting on a vector (v) if the matrix components are the probability to substituting one unit for another one, and if the vector components represent the probabilities of having a given unit in the configuration. The probability of some configuration of n base-pairs (bps) is modeled by n successive applications of the matrix M on an initial vector $v_0$, which characterizes the initial condition on the system. After n applications, the final configuration of the system can be written as a linear combination of the eigenvectors associated with M, i.e., $v_n = \sum_{m=1}^{\infty} a_m \lambda_m^n e_m$, where $e_m$ are the eigenvectors M with eigenvalues $\lambda_m^n$ and $a_m$ are the projections of $v_0$ into $e_m$.

A matrix with all the columns normalized to one, as $M^a$, has the property that at least one eigenvalue is one, while the real part of all others eigenvalues 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 on $v_0$. If we assume that M has one eigenvector $e_1$ with eigenvalue equal to one, then in the limit of large n, $v_0^n$ converges to $e_1$, with $a_1 = 1$, due to 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

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

Solving Eq. (1), the probability of any configuration in the system is obtained.

The NN model for nucleic acids assumes that the stability of a given bp depends on the identity and orientation of the neighboring base pair. The application of the NN model to nucleic acid was pioneered by Zimm [21] and by Tinoco and coworkers [22–26]. For our theoretical description, the NN model allows us to construct the stochastic matrix that describes DNA.

To construct the stochastic matrix that allow us to describe the melting behavior of DNA, we first need to define the units. These units must be given by two combinations: ...Poly(dpyrimidine). The four sites give 16 different combination of base pairing. For DNA the units are covalently bonded by a phosphodiester from its 5'-hydroxyl group to the 3'-hydroxyl group of one neighbor and by a second phosphodiester from its 3'-hydroxyl to the 5'-hydroxyl of its other neighbor.

The 16 different combinations for neighbors can be displayed as a $4 \times 4$ matrix, namely

$$
\begin{pmatrix}
\uparrow \text{Pu} \cdot \text{Py} & \downarrow \text{Py} \cdot \text{Pu} & \uparrow \text{Pu} \cdot \text{Py} & \downarrow \text{Py} \cdot \text{Pu} \\
\uparrow \text{Py} \cdot \text{Pu} & \downarrow \text{Pu} \cdot \text{Py} & \uparrow \text{Py} \cdot \text{Pu} & \downarrow \text{Pu} \cdot \text{Py} \\
\uparrow \text{Pu} \cdot \text{Py} & \downarrow \text{Py} \cdot \text{Pu} & \uparrow \text{Pu} \cdot \text{Py} & \downarrow \text{Py} \cdot \text{Pu} \\
\uparrow \text{Py} \cdot \text{Pu} & \downarrow \text{Pu} \cdot \text{Py} & \uparrow \text{Py} \cdot \text{Pu} & \downarrow \text{Pu} \cdot \text{Py} \\
\end{pmatrix}
$$

(2)
where \( \uparrow \text{Pu} \cdot \downarrow \text{Py} \) represents the probability of a bonding \( \uparrow \text{Pu} \cdot \downarrow \text{Py} \downarrow \), \( \uparrow \text{Pu} \cdot \downarrow \text{Py} \) represents the probability of a bonding \( \uparrow \text{Pu} \cdot \downarrow \text{Py} \downarrow \) neighboring an nonbonding \( \uparrow \text{Pu} \downarrow \text{Py} \), etc. These stacking are in three dimensions and this information must be included in the stacking energy. 

Each bp stacking has a finite probability of occurrence. The statistical weight for each process is one, because only one form of stacking exist. Based on the NN model each configuration is proportional to its stability constant \( s_i = \Delta G_i - T \Delta S_i \) where \( i = \text{PuPy} \) or \( \text{PyPu} \). Inserting all the energetic contributions in matrix (2), the explicit matrix is written as

\[
\begin{pmatrix}
e^{-\frac{\Delta G_{\text{PuPu}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PuPy}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PyPu}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PyPy}}}{k_B T}} \\
e^{-\frac{\Delta G_{\text{PuPy}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PyPy}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PyPu}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PuPu}}}{k_B T}} \\
e^{-\frac{\Delta G_{\text{PyPu}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PyPy}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PuPu}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PyPu}}}{k_B T}} \\
e^{-\frac{\Delta G_{\text{PyPy}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PuPu}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PyPu}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PyPy}}}{k_B T}}
\end{pmatrix},
\]

where \( \Delta G_{\text{MN}} \) is the free energy between the \( \uparrow M \cdot N \downarrow \) nearest-neighbor, \( \Delta G_h \) is the free energy due to hydrogen bonds, \( k_B \) is the Boltzmann constant and \( T \) the temperature.

For matrix (3), the eigenvector with eigenvalue equal to one is a vector with four components that gives us the probability of finding any of the following configurations in the system \( \uparrow \text{Pu} \cdot \downarrow \text{Py} \downarrow, \uparrow \text{Pu} \cdot \downarrow \text{Py} \downarrow, \uparrow \text{PuPy} \downarrow \) or \( \uparrow \text{PyPu} \downarrow \). The sum \( \uparrow \text{PuPy} \downarrow + \uparrow \text{PyPu} \downarrow \) gives us the probability of having the denatured bps in the system. By setting the denatured probability to \( 1/2 \), \( T_m \) is obtained for DNA chains for large \( n \). Large \( n \) means chains larger than 20 bps and shorter than 350 bps. The lower limit is imposed because below 20 iterations of \( v \) applied to \( M \), the probability is not stable. The upper limit is based on the length of cooperatively melting regions which are 350 bps. In our model, we do not take into account the influence of cooperatively melting regions in the melting process. Although we do not take into account in our theoretical description of denaturation, one of the interesting characteristics of biological macromolecules is the interplay between the cooperative interactions among regions and the independent properties of these regions [4].

In the next section we use matrix (2) to derive a theoretical expression that allows us to calculate the \( T_m \) for two simple systems, poly(dPu)-poly(dPy) and poly(dPu-dPy)-poly(dPu-dPy).

3. Results and comparison with experiment

In this section we use matrix (2) to calculate \( T_m \) for poly(dPu)-poly(dPy) and poly(dPu-dPy)-poly(dPu-dPy). To obtain the \( T_m \) we proceed as follows: obtaining the eigenvector with eigenvalue equal to one we can find the probability to have natured and denatured bps. Using the definition of \( T_m \) we set the probability of denatured bps equal to \( 1/2 \) and solving the equation for \( T_m \), we can find the melting temperature.

3.1. Calculation of \( T_m \) for poly(dpurines)-poly(dpyrimidines)

For this particular case only the terms \( \uparrow \text{Pu} \downarrow \) are conserved in matrix (2), and this matrix is reduced to \( 2 \times 2 \) matrix, namely

\[
\begin{pmatrix}
\uparrow \text{Pu} \cdot \downarrow \text{Py} \\
\uparrow \text{Py} \cdot \downarrow \text{Pu} \\
\uparrow \text{PuPy} \\
\uparrow \text{PyPu}
\end{pmatrix}
\]

(4)

Inserting the energetic contributions in matrix (4) we find that,

\[
\begin{pmatrix}
e^{-\frac{\Delta G_{\text{PuPu}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PyPy}}}{k_B T}} \\
e^{-\frac{\Delta G_{\text{PuPy}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PyPu}}}{k_B T}} \\
e^{-\frac{\Delta G_{\text{PuPu}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PyPy}}}{k_B T}} \\
e^{-\frac{\Delta G_{\text{PyPu}}}{k_B T}} & e^{-\frac{\Delta G_{\text{PyPy}}}{k_B T}}
\end{pmatrix},
\]

where \( \Delta G_{\text{PuPu}} \) is the free energy for the dimer \( \uparrow \text{Pu} \cdot \downarrow \text{Py} \) and \( \Delta G_{\text{PyPu}} \) is the free energy for the hydrogen bond between Pu and Py. If the hydrogen bonds of a bp are broken we have \( \Delta G_{\text{PuPu}} = \Delta G_{\text{PyPu}} \), and if the hydrogen bonds of a dimer are broken we have \( \Delta G_{\text{PuPu}} = 2\Delta G_{\text{PyPu}} \).

After normalizing each column of matrix (5), we obtain
The explicit form of the eigenvector with eigenvalue one is obtained solving Eq. (1), which for the present case yields the following vector,

\[
\begin{pmatrix}
\frac{1}{1 + e^{-\frac{\Delta G_{h_{PuPy}}}{k_B T}}}
\frac{\Delta G_{h_{PuPy}}}{k_B T}
\frac{1}{1 + e^{-\frac{\Delta G_{h_{PuPy}}}{k_B T}}}
\end{pmatrix}
\]

(7)

Vector (7) gives us the probability of finding \( \uparrow \text{Pu} \cdot \downarrow \text{Py} \) and \( \uparrow \text{Pu} \downarrow \downarrow \text{Py} \) base-pairs in poly(dPu)-poly(dPy) at any temperature. Now, if we want to calculate the \( T_m \) we only need to set the second component of vector (7) equal to one half,

\[
1 - \frac{1}{1+e^{-\frac{\Delta G_{h_{PuPy}}}{k_B T}}} = \frac{1}{2}
\]

(8)

Equation (8) implies that \( \Delta G_h = 0 \), and this is the condition for finding the melting temperature. \( \Delta G_{h_{PuPy}} \) can also be calculated from \( \Delta H_{h_{PuPy}} \) and \( \Delta S_{h_{PuPy}} \) by using the equation

\[
\Delta G_{h_{PuPy}} = \Delta H_{h_{PuPy}} - T_m \Delta S_{h_{PuPy}},
\]

(9)

where \( \Delta H_{h_{PuPy}} \) is the enthalpy due to a hydrogen bonding and \( \Delta S_{h_{PuPy}} \) its respective entropy.

If \( \Delta G_{h_{PuPy}} = 0 \), then

\[
T_m = \frac{\Delta H_{h_{GC}}}{\Delta S_{h_{GC}}}
\]

(10)

It is very important to remark that Eq. (10) depends only on hydrogen bond parameters and has no influence on the stacking free energies, as was shown by Wartell and Be-night [4].

If we want to calculate \( T_m \) for poly(dG)-poly(dC), guanine is the purine, and cytosine is the pyrimidine. Experimentally, \( \Delta H_{h_{GC}} = -5.8 \text{ kcal/mol} \) [28] and \( \Delta S_{h_{GC}} = -16 \text{ e.u.} \). Inserting these values in Eq. (10), we found that the melting temperature for poly(dG)-poly(dC) is 362.35 K, compared to the value of experiment 360.8 K [29], only 1.5 K difference. Also Eq. (10) implies that the process is given in one step; this means that the entire chain is denatured at once at \( T_m \) (two-state melting transition).

Also, Eq. (10) can be used to calculate \( \Delta S \) since the experimental values for the melting temperature and \( \Delta H \) are more commonly found in the literature. For poly(dA)-poly(dT) these values are \(-4 \text{ kcal/mol} \) and 53 K respectively [36]. Using these values in Eq. (10), we obtain \( \Delta S_{h_{AT}} = -12.26 \text{ e.u.} \); the experimental values are between \(-11 \text{ e.u.} \) and \(-13 \text{ e.u.} \) [36].

3.2. Calculation of \( T_m \) for poly(purine-pyrimidine)-poly(purine-pyrimidine)

In this subsection using matrix (2) we find the \( T_m \) for poly(dPu-dPy)-poly(dPu-dPy).

For this particular case, in matrix (2) only the terms \( \uparrow \text{Pu} \downarrow \downarrow \text{Py} \) are conserved, obtaining

\[
\begin{pmatrix}
0 & \downarrow \text{Py} \cdot \downarrow \text{Pu} & \uparrow \text{Py} \cdot \uparrow \text{Pu} & 0 & \downarrow \text{Py} \cdot \downarrow \text{Pu} & \uparrow \text{Py} \cdot \uparrow \text{Pu} & 0 & \downarrow \text{Py} \cdot \downarrow \text{Pu} & \uparrow \text{Py} \cdot \uparrow \text{Pu} & 0 \end{pmatrix}
\]

(11)

Inserting the energetic contributions in the last matrix we obtain,

\[
\begin{pmatrix}
0 & e^{-\frac{\Delta G_{pyPy}}{k_B T}} & e^{-\frac{\Delta G_{puPu}}{k_B T}} & 0 & e^{-\frac{\Delta G_{pyPy}}{k_B T}} & e^{-\frac{\Delta G_{puPu}}{k_B T}} & 0 & e^{-\frac{\Delta G_{pyPy}}{k_B T}} & e^{-\frac{\Delta G_{puPu}}{k_B T}} & 0 \end{pmatrix}
\]

(12)

After normalizing each column of matrix (12), solving Eq. (1) with \( \mathbf{M} \) given by Eq. (12) and adding \( \uparrow \text{PuPy} \downarrow + \uparrow \text{PuPy} \downarrow = 1/2 \), we find the following condition to obtain the \( T_m \)

Equation (13) depends on the hydrogen parameters as the staking ones. In this case the stacking energies play a fundamental roll, and the competition between $\Delta S$’s and $\Delta H$’s govern the melting behavior.

We can expand Eq. (13) in a Taylor series around $T_m$ and taking terms up to first order we find

$$\Delta G_{PuPy} + \Delta G_{PyPu} = 0. \quad (14)$$

Inserting $\Delta G_{PuPy} = \Delta H_{PuPy} - T_m \Delta S_{PuPy}$ and $\Delta G_{PyPu} = \Delta H_{PyPu} - T_m \Delta S_{PyPu}$ in Eq. (14) and solving to $T_m$, we obtain,

$$T_m = \frac{\Delta H_{PuPy}}{\Delta S_{PuPy}} + \frac{\Delta H_{PyPu}}{\Delta S_{PyPu}}. \quad (15)$$

If $\Delta S_{PuPy} \approx \Delta S_{PyPu}$, which is well accepted for polynucleotides [36], finally, we find

$$T_m \approx \frac{\Delta H_{PuPy} + \Delta H_{PyPu}}{\Delta S} \approx T_{PuPy}^m + T_{PyPu}^m. \quad (16)$$

Equation (16) reproduce the well known result $T_m = \sum T_{MN}$. This last Equation is the most often used to obtain the $T_m$, starting from the nearest-neighbors thermodynamics.

It is difficult to decide which experimental parameters to use in Eqs. (13) or (16), because there has been disagreement concerning this issue: particularly the difference between DNA polymers and oligonucleotides, MN thermodynamic trends and the salt dependence on nucleic denaturation. The major source of confusion in the literature is that the different studies use different oligonucleotide and polymer design, different methods for determining thermodynamics, different method for analyzing data, different salt conditions, and different formats to present MN parameters. Some of the experimental values of nearest-neighbor thermodynamics can be consulted in Refs. [29–35]. Also theoretical efforts have been spent to calculate stacking energies by \textit{ab initio} calculations, for a review consult reference [5].

Even though it is difficult to decide which set of parameters to use in Eqs. (13) or (16) to calculate $T_m$, we introduce the hydrogen parameters obtained by Newmark [28] and the nearest-neighbors thermodynamics values obtained by Delcourt [32] to obtain $T_m$ for poly(dG-dC)-poly(dG-dC). The $T_m$ predicted by our method using Eq. 13 is 99.47° C while experimentally it is 99.2° C [29].

Also Eq. (13) could help to discern which set of data is the best one, since it is only well established that $\uparrow Pu \cdot Py \downarrow$ is more stable than $\uparrow Py \cdot Pu \downarrow$ [36].

4. Conclusions

In this article we present a theoretical framework that allows us to predict $T_m$ of poly(dG)-poly(dC) and poly(dG-dC) poly(dG-dC) without adjustable parameters. This theoretical framework was supported in the description of denaturation of DNA by the stochastic matrix method.

For poly(dpurine)-poly(dpyrimidine) we obtain an interesting result for the melting temperature: it only depends on $\Delta G_{PuPy}$ (the free energy for the hydrogen bonds between purines and pyrimidines), and at $T_m$ it is found that $\Delta G_{PuPy} = 0$. Using the relation

$$\Delta G_{PuPy} = \Delta H_{PuPy} - T_m \Delta S_{PuPy} = 0,$$

we obtain $T_m = \Delta H_{PuPy}/\Delta S_{PuPy}$. Is important to remark that this last result does not depend on the stacking energies and predicts that the denaturation of this chain occurs in one single step.

For poly(dpurine-dpyrimidine)-poly(dpurine-dpyrimidine) we find the well known relation $T_m = \sum T_{MN}$ but only under specific restrictions between stacking energies and entropies. On one hand the ratio between stacking energies and $k_B T$ must be appropriate to apply a Taylor expansion, and on the other, hand $\Delta S_{PuPy} \approx \Delta S_{PyPu}$.

One of the most important applications of our method could be the systematic study of the melting transition changing the length and the composition of the DNA.

Finally, it is important to remark that in this paper we give a theoretical framework that allows us to obtain a theoretical description of DNA denaturation and to calculate its melting temperature without adjustable parameters. In a future work this description could be extended, taking into account the nature of purines and pyrimidines, \textit{i.e.}, to include adenines, guanines, thymines and cytosines in the stochastic matrix. Also, our method could be used in the study of denatured proteins.

Acknowledgements

L.D. gratefully acknowledges the guidance and useful conversations with Professor L.S. García-Colín who contributed the best comprehension of some of the ideas exposed in this work. Also authors acknowledge with gratitude the important conversations with Germinal Cocho and O. Resendiz-Antonio at the beginning of this work. The authors want to thank Stephie and M.V. Vázquez for help in preparation of the english version of this article. Finally, we are in debt with Lynn Young for helpful suggestions and critical reading of the manuscript.

Arrows designate the direction of the sugar-phosphate chain, from C3′ atom of a deoxyribose unit to C5′ atom of the next deoxyribose adjacent to and on either side of the phosphodiester linkage. Some times nearest/neighbor base pairs are represented with a slash separating strands in antiparallel orientation (e.g., AC/TG means 5′−AC−3′ Watson-Crick base paired with 3′−TG−5′ or ↓A·T↑ C·G↓ in the notation used throughout this paper).

Abbreviations: e.u., entropy units (cal/Kmol).

It should be stressed that the calculated energies by the ab initio method consider only molecules in vacuo and do not take into account hydrophobic interactions which, however, contribute significantly to stacking interactions.