Volume 13 Number 16 1985 Nucleic Acids Research

Melting transition of covalently closed DNA with supercoil-induced cruciforms

Rabi Majumdar* and Ashoke R.Thakur+

Saha Institute of Nuclear Physics, and +Department of Physics, University of Calcutta, 92, A.P.C. Road, Calcutta 700009, India

Received 20 June 1985; Accepted 5 August 1985

ABSTRACT

The melting curve for covalently closed supercoiled DNA has been studied by assuming the existence of cruciforms as significant structural perturbations in the pre-melting region. The statistical mechanical treatment used incorporates these cruciform structures through an appropriate sequence generating function. The variation of the effective hydrogen bond energy with temperature is taken into account by an empirical procedure. The results obtained are in close agreement with the corresponding experimental data in TEA solution where the effect of heterogeneity of the base pairs is minimized.

INTRODUCTION

The existence of inverted repeat sequences is known to be quite common in DNA molecules. Although their functional significance is not yet well established, it is often believed that the self-complementarity of these inverted repeats in DNA provides the structural intermediates for mutation[1]. In principle, the segment of a DNA molecule with an inverted repeat has two distinct potentially stable conformations, depending on whether the hydrogen bonds are formed between the two separate strands or within each single strand of the sequence. These are the usual uninterrupted double helical structure and the 'cruciform' structure respectively. The cruciform structures are found to exist at certain inverted repeats within the various covalently closed plasmids[2,3] as well as in a completely palindromic circular DNA molecule[4]. It is important to note, however, that the cruciform states are observed only in covalently closed circular duplex structures (DNA I) with negative superhelix density, but never in a nicked circular duplex (DNA II) or in a linear molecule. Thus for the equilibrium secondary structure of the

underwound DNA I, the cruciform state appears to be thermodynamically stable under the appropriate strain of supercoiling[4,5]. In other words, the excess free energy of supercoiling, in such cases, must compensate for the disruption of the base pairing and stacking needed to drive the cruciform extrusions. For a closed circular molecule, we know that this free energy is proportional to the square of the superhelix density[6-8]. Moreover, since a stronger preference for the cruciform structure has been reported at higher temperatures[4], we find it reasonable to assume that the DNA I, beyond some critical value of supercoiling,will largely relax into a set of cruciforms in the pre-melting region[9]. In other words, the present model for DNA I, in the transient region, will consist of three types of sequences, namely those consisting of the regular helical units, the denatured regions or the bubbles, and the supercoil-induced cruciforms. The DNA II, on the other hand, will be represented by a linear duplex consisting of the regular helical sequences and the bubbles only. Thus we consider the presence of the supercoil-induced cruciforms as significant structural perturbations in DNA I in analyzing its melting curve. The states of partial cruciform extrusion, however, are not considered in this paper. We have always ignored the denatured loops at the heads of an extruded cruciform as these loops are at an entropically disadvantageous position to expand beyond their minimum size. Strictly speaking, however, this is true for a palindromic sequence only.

A statistical mechanical treatment of superhelical cruciform extrusion has appeared recently[10]. In the present paper, however, we use the appropriate sequence-generating functions[11,12] for computing the equilibrium transition properties. In fact the main result of this paper indicates that the higher melting point and the flatter melting curve for DNA I, as observed recently[13], can be well explained by introducing into the usual statistical mechanical theory an additional sequence generating function for the supercoil-induced cruciforms, and by considering the fact that the process of nucleation is easier in a closed duplex structure. The order of magnitude of the relevant nucleation parameter, which determines the coope-

rativity of the order-disorder transition, is also discussed. The effect of heterogeneity of the base pairs, however, is not considered as we have compared our results with the experimental data on melting profiles in tetraethylammonium bromide (TEA) solution where the difference between the A-T and the G-C binding energies are minimized[13]. Finally, in order to obtain a closer fit to the observed melting profiles over a wider temperature range, we have also taken into account the variation of the effective hydrogen bond energy with temperature, and some important conclusions related to this variation are summarized at the end.

THEORY

The thermodynamics of melting of the hydrogen bonded polymers, with special reference to the linear duplex polynucleotide molecules (DNA II), has been worked out by several authors on the basis of Ising-like models in statistical mechanics[14]. The cooperativity of the hydrogen bonds, which brings out the sharp melting transition for a linear duplex molecule with high degree of polymerization, has been discussed by one of us[15] in a recent review. As our present work includes the closed circular duplex molecules (DNA I) as well, we find it convenient to use the sequence-generating functions, first oroposed by Lifson[11], in order to obtain the appropriate weight factors or the partition function. This method, although basically similar to the usual matrix method in an Ising-like model, has a wider range of applicability and can be suitably extended to include the supercoil-induced cruciforms in DNA I.

It is reasonable to assume that the free-end effects of a sufficiently long linear duplex molecule are unimportant, so that a long-chain DNA II molecule in the pre-melting region effectively consists of the ordered helical (h) sequences and the bubbles only. Thus the diordered coiled (c) sequences are assumed to exist only in the form of bubbles. The corresponding sequence-generating functions are given by[11,12]

$$h(x) = \sigma_n \sum_{i=1}^{\infty} \left(\frac{s}{x}\right)^i = \sigma_n s/(x-s) \qquad (1)$$

$$c(x) = \sum_{i=1}^{\infty} \left(\frac{1}{x}\right)^i i^{-k} \qquad (2)$$

for the helical and the coiled sequences respectively, where s represents the 'equilibrium parameter' for the hydrogen bond formation reaction in a link that follows a bonded link, and σ_n is the 'nucleation parameter' which mainly depends on the 'stacking energy' for a duplex polynucleotide molecule[14,15]. We know that $\sigma_n \ll 1$ imposes high degree of cooperativity of the hydrogen bonds. The nature of the melting transition is also determined by the constant k, which is related to the conformational degrees of freedom of the bubble due to its flexible bonds. According to a crude estimate[12] based on the random walk model in three dimensions, $k \simeq 1.5$. The effect of variation of the constant k has been discussed in an earlier paper[16], where it was actually pointed out that $k < 1.5$ for a closed circular molecule. This is because the full degrees of freedom of a bubble cannot be realized in a closed duplex structure, where the flexibility of the strands is constrained by a fixed linking number. In our present work, we take k = 1, so that the summation in equation (2) may be evaluated exactly. Thus we have

$$\sum_{i=1}^{\infty} \left(\frac{1}{x}\right)^i i^{-1} = -\ln\left(1 - \frac{1}{x}\right) \qquad (3)$$

Now, for a linear chain of n units, each of which can exist in the helical or in the disordered state, Lifson[11] has shown that the partition function, in the limit of large n , is given by

$$Z = x^n \qquad (4)$$

where x is the largest root of the equation

$$h(x)\, c(x) = 1 \qquad (5)$$

Combining equations (1), (2), (3) and (5), we get

$$(x-s)/\sigma_n s = -\ln\left(1 - \frac{1}{x}\right) \qquad (6)$$

where it is reasonable to choose $\sigma_n = 0.01$ for the linear duplex structure[17]. The equation (6) can be solved graphically for x at different values of s referring to different tem-

peratures. The fraction of the helical states is then obtained from the relation[12]

$$\theta = \frac{s}{x} \cdot \frac{dx}{ds} \qquad (7)$$

Finally, using equation (6) to evaluate the derivative, we get

$$\theta = \left[1 + \sigma_n s/x(x-1) \right]^{-1} \qquad (8)$$

where the parameter s is translated into the absolute temperature T by using the relation[14]

$$s = \exp\left(-\Delta E_B/RT\right) \cdot \exp\left(\Delta S/R\right) \qquad (9)$$

In equation (9), E_B represents the hydrogen bond energy, and S is the entropy of a participating link of the polymer arising out of its conformational degrees of freedom.

As our present paper is aimed at interpreting the observed melting curve for DNA I, let us now turn our attention to the covalently closed duplex polynucleotide with negative superhelix density. As mentioned already, we find it reasonable to assume that the negative supercoiling largely manifests itself in a state with cruciforms in the pre-melting region. Thus, in the present model, the DNA I effectively consists of three types of sequences, namely those with the regular helical states (h), the coiled states (c), and the helical states (h^+) in the supercoil-induced cruciforms. An inspection of the cruciform structure leads us to express the sequence-generating function for the h^+-sequences in the following form:

$$h^+(x) = \sigma_n^* \sum_{i=1}^{\infty} \left(\frac{s^2}{x^4}\right)^i = \sigma_n^* s^2/(x^4 - s^2) \qquad (10)$$

Equation (10) reflects the fact that the helical sequences in a cruciform should grow by four bases at a time, and that s is still the statistical weight per base pair, or $s^{1/2}$ per base. Here we find it more reasonable to consider the growth of the helical sequences by four bases at a time, rather than by two bases at a time as if the pair of hairpin structures in a cruciform were growing independently. The factor σ_n^* involves the energy necessary to nucleate a cruciform structure, following a regular helical sequence, causing the disruption of a minimum number of base pairs. These local transi-

tions are rendered energetically favourable by the imposed superhelicity in closed circular molecules. Therefore, the factor σ_n^* actually involves the 'average' superhelical free energy cost of a cruciform extrusion in such molecules. The superhelical free energy is estimated to be $\sim 10RT\rho^2$ per base-pair[6-8], where the superhelix density $\rho \simeq -0.05$ to -0.06 for the commonly occurring circular forms[18]. As the average number of base pairs involved in a cruciform extrusion is only a few, we take $\sigma_n^* \simeq 1$. Thus, for the sufficiently lower values of ρ involved, we actually ignore the local dependence of σ_n^* on superhelicity, but the structural manifestations of the latter are taken into account by equation (10). For the regular helical sequence in a supercoiled molecule, on the other hand, the equation (1) is still valid with a modified nucleation parameter $\sigma_n' > \sigma_n$. This is quite reasonable as the nucleation of a helical sequence should be easier in a covalently closed structure where the conformations are more restricted by the topological constraints. The relative magnitudes of these parameters are discussed later. Thus, in the present case, we have

$$h'(x) = \sigma_n' s/(x-s) \qquad (11)$$

For the coiled sequence, however, we assume that the equation (2) remains the same, so that the partition function Z is still given by equation (4) where, according to Lifson[11], x is now the largest root of the determinantal equation

$$\left| M_{ij}(x) - \delta_{ij} \right| = 0 \qquad (12)$$

where

$$M_{ij}(x) = \begin{vmatrix} 0 & h'(x) & 0 \\ 0 & 0 & h^+(x) \\ c(x) & 0 & 0 \end{vmatrix} \qquad (13)$$

and i, j = 1,2,3 represent the coiled, the helical and the cruciform sequences respectively. As the superhelical cruciform extrusions are entropically favourable in the helical

regions, we have assumed a regular order for the three types of sequences involved. Thus, combining equations (12) and (13), we get

$$h'(x)\, h^{+}(x)\, c(x) = 1 \tag{14}$$

or, using equations (2), (3), (10) and (11), we finally obtain

$$(x^4 - s^2)(x-s)/\sigma_n^{+} s^3 = -\ln\left(1 - \frac{1}{x}\right) \tag{15}$$

where $\sigma_n^{+} \simeq \sigma_n' \sigma_n^{*}$ should be treated as a new effective nucleation parameter for the supercoiled DNA. Now, we have to solve the equation (15) graphically for its largest root x, like our previous equation (6) for the linear system, for the different values of s which refer to different temperatures T according to equation (9). The fraction of the helical states at these temperatures may then be obtained from equation (7). Using equation (15) to evaluate the derivative dx/ds, in the present case, we obtain

$$\theta = \frac{\frac{1}{s} + \frac{2x^3}{s^2} - \frac{3x^4}{s^3}}{\frac{1}{s} + \frac{4x^3}{s^2} - \frac{5x^4}{s^3} - \frac{\sigma_n^{+}}{x(x-1)}} \tag{16}$$

We have finally plotted the $(1-\theta)$ vs. T melting curve for DNA I as well as that for DNA II by taking $\sigma_n^{+} = 0.1$ and $\sigma_n = 0.01$ respectively, and by choosing the appropriate value of the hydrogen bond energy E_B and the entropy S appearing in equation (9). The results are compared with the corresponding experimental data on the melting profiles in TEA solution where the effect of heterogeneity of the base pairs is minimized[13]. The physical basis for the choice of parameters and the nature of agreement between theory and experiment are discussed below.

RESULTS AND DISCUSSIONS

The main aim of this paper is to understand the nature of the observed melting profile for DNA I by assuming that the free energy associated with supercoiling is manifested, through structural reorganization, as a series of cruciform sequences formed within the covalently closed duplex polynucleotide. Accordingly, we have solved the equation (15) gra-

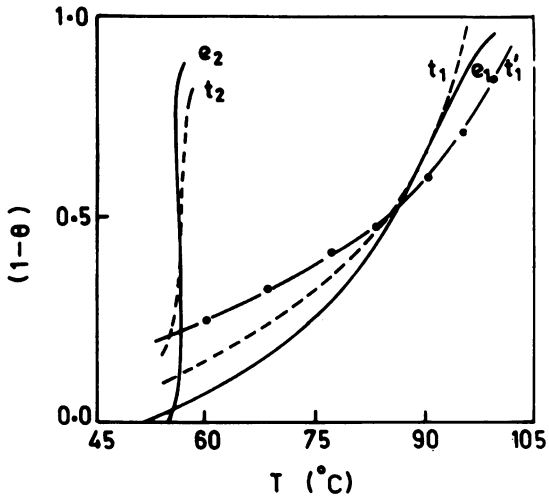

Fig.1. The Melting Curves for DNA I and DNA II. The experimentally observed curve, e_1 and e_2 refer to DNA I and DNA II respectively. The theoretically computed curves t'_1, t_1 refer to DNA I and t_2 refers to DNA II.

phically, and computed the fractions of the helical states from equation (16) on the basis of an effective nucleation parameter σ_n^+ for the supercoiled DNA. The corresponding temperature T are calculated from equation (9) which involves the hydrogen bond energy E_B and the entropy S . In obtaining the melting profile of DNA II on the other hand, we have solved the equation (6) graphically, with $\sigma_n = 0.01$, as in ref.17. The main results of our paper are summarized below.

(a) In Fig.1, the curve t'_1 shows the theoretically computed melting profile for DNA I with $\sigma_n^+ = 0.1$ and with E_B=4000 cal/mole and S_B=11.53 e.u. As mentioned before, we expect the nucleation of a helical sequence to be easier in a covalently closed system, implying reduced base-pair cooperativity, so that $\sigma_n^+ > \sigma_n$. The choice $\sigma_n^+ = 0.1$, in the present case, reproduces the observed slope of the experimental melting curve e_1. For E_B= 4000 cal/mole, the observed melting point T_m= 84 .0°C for DNA I, on the other hand, is reproduced by the choice S = 11.53 e.u. It may be remarked that this choice

is reasonably close to the corresponding estimated value in our earlier paper[16].

(b) Having chosen the parameters σ_n^+, E_B and S above, we find that the agreement between the curves t_1' and e_1 in Fig.1, specially in the lower temperature region, is still not quite satisfactory. However, it is interesting to note that our results may be improved considerably, as shown by the theoretical curve t_1, if we assume that E_B varies linearly from 4200cal/mole to 3900 cal/mole in the temperature range between 57° C and 95° C. Thus we find that an effective variation of the hydrogen bonded interaction, between the two strands, with temperature may be considered reasonable, as pointed out by several others[19,20].

(c) Extrapolating the value of E_B to the observed melting point T_m = 57.0° C for DNA II, we obtain E_B=4200 cal/mole. The corresponding choice S = 12.80 e.u. for DNA II, with σ_n = 0.01, as mentioned before, reproduces the observed melting point of the molecule. The calculated melting profile t_2 also gives a good fit to its observed profile e_2.

(d) From the phenomenological considerations described above, we clearly find that S = 11.53 e.u. for DNA I, while S = 12.80 for DNA II. Thus, we are perfectly justified in assuming that the process of nucleation is easier in a covalently closed molecule, where the topological constraint, due to fixed linking number, leads to a restriction in its conformational degrees of freedom, causing loss of entropy. Thus, from our present considerations, we are able to obtain a quantitative estimate for this entropy loss. It has often been suggested that the supercoiling of DNA is enthalpy-determined[18], but it is interesting to note that our estimated entropy loss helps us to understand the relative magnitudes of the nucleation parameters σ_n and σ_n', and hence the value σ_n' = 0.1 used by us. Clearly, the entropy considerations alone should contribute a relative factor of $\exp(\delta S/R) \sim 2$ in the present case. At the same time, it should also be noted that the stacking of bases is generally less disturbed by the helix-coil transitions in a covalently closed molecule due to the conformational restrictions involved. Since the

energy factor, arising out of such considerations in the present case, is known to be more important than the entropy factor[17,18], the former is expected to contribute a relative factor $>$ 2-3. This explains the net relative factor $\sim$10 between σ_n and σ_n'. Since $\sigma_n^+ \simeq \sigma_n'\sigma_n^*$, and the parameter σ_n^* appearing in equation (10) is relatively closer to unity, the relative magnitudes of σ_n and σ_n^+ used by us, namely σ_n = 0.01 and σ_n^+ = 0.1, are well justified. It may be pointed out that a similar variation in the degree of base-pair cooperativity, depending on the constraints involved, has been examined recently by others[21] in connection with the melting of short linear DNA fragments.

Thus, we are able to explain the main features of the melting curve for DNA by assuming that the free energy of supercoiling is manifested, through structural reorganization, as a series of cruciform sequences. The dynamics of these supercoil-induced cruciforms, introduced into our theory through a new sequence-generating function, explains the observed higher melting point of DNA I, as compared to that of DNA II. The flatter melting curve for DNA I, which implies lack of cooperativity of its hydrogen bonds , is explained by the easier nucleation property, namely $\sigma_n^+ > \sigma_n$. It should be remembered, however, that our results are compared with the corresponding data on the melting profiles in TEA solution[13] where the effect of heterogeneity of the base pairs is minimized.

ACKNOWLEDGEMENTS

The authors wish to thank Prof. B. Dutta-Roy and Dr. D. Bhowmik for having the benefit of very useful discussions with them during the course of this work.

*To whom correspondence should be addressed

REFERENCES

1. Ripley, L. S. and Glickman, B. W. (1983) Cold Spring Harbor Symp.Quant.Biol. 47, 851-861.
2. Lilley, D. M. J. (1980) Proc.Natt.Acad.Sci. (U.S.A.) 77, 6468-6472.

3. Panayatatos, N. and Wells, R.D. (1981) Nature (Lond.) 289, 466-470.
4. Mizuuchi, K., Mizuuchi, M. and Gellert, M. (1982) J.Mol.Biol.156, 229-243.
5. Singleton, C. (1983) J. Biol.Chem. 258, 7661-7668.
6. Hsieh, T. H. and Wang, J. C. (1975) Biochemistry 14, 527-535.
7. Depew, R. E. and Wang, J.C. (1975) Proc.Natl.Acad.Sci. (U.S.A.) 72, 4275-4279.
8. Pulleyblank, D. E., Shure, M., Tang, D., Vinograd, J. and Vosberg, H.P., (1975) Proc.Natl.Acad.Sci.(U.S.A.), 72, 4280-4284.
9. Vologodskii, A. V., Lukashin, A.V., Anshelevich, V.V. and Frank-Kamenerskii, M.D. (1979) Nucleic Acids Res. 6,967-982.
10. Benham, C.J. (1982) Biopolymers 21, 679-696.
11. Lifson, S., (1964) J.Chem.Phys. 40, 3705-3710.
12. Poland, D. and Scheraga, H. (1964) J.Chem.Phys. 45, 1464-1469.
13. Gauga, A. V., Belintsev, B.N. and Lyubchenko, Yu.L.(1981) Nature 294, 662-663.
14. Poland, D. and Scheraga, H. (1970) Theory of Helix-Coil Transitions in Biopolymers, Academic Press, p.1-322.
15. Majumdar, R. and Pathria, R.K. (1985) J. Macromol.Sci. C (in press).
16. Majumdar, R. and Thakur, A. R. (1985) Int.J. Quantum Chem. Quantum Biol.Symp. 12 (in press).
17. Zimm. B. H. (1960) J. Chem. Phys. 33, 1349-1356.
18. Seidl, A. and Hinz, H-J (1984) Proc.Natl.Acad.Sci. (U.S.A.) 81, 1312-1316.
19. Luck, W. A. P. (1976) In 'The Hydrogen Bond III' Edited by P. Schuster, G. Zundel and C. Sandorfy, North-Holland Publishing Co. p.1367-1423.
20. Ghosh, P.N. (1977) J. Phys. C10, 4421-4424.
21. Benight, A. S. and Wartell, R. M. (1983) Biopolymers 22, 1409-1425.