

## Applicability of sequence generating techniques in supercoiled DNA

Ansuman Lahiri

Theoretical Nuclear Physics Division, Saha Institute of Nuclear Physics,  
92 Acharya Prafulla Chandra Road, Calcutta -700 009, India

**Abstract :** The applicability of sequence generating techniques to supercoiled DNA has been shown to depend on the functional form of the supercoiling energy. For partially melted closed circular DNA a range of superhelical densities may be found out where these techniques become applicable. These observations may have implications for partition function calculation of supercoiled DNA with arbitrary base sequence and composition.

**Keywords :** Sequence generating techniques, supercoiled DNA, helix-coil transition.

**PACS Nos :** 87.15.By, 36.20.Ey

### 1. Introduction

Physicochemical studies of supercoiled DNA which involve the calculation of long chain partition functions have so far approached the problem from the point of view of the combinatorial method

$$Z = \sum g e^{-\beta E} \quad (1)$$

where  $Z$  is the configuration at partition function, the sum in the right hand side being over all possible configurations of the system concerned. Here  $g$  denotes the degeneracy of the configuration,  $\beta = 1/kT$ , where  $k$  is the Boltzmann constant and  $T$  is the absolute temperature, and  $E$  is the configurational energy of the system.

In general, the formulation of the partition function in terms of explicit degeneracy factors is difficult for complicated systems since the corresponding combinatorial problem tends to become intractable. So far, only a few problems have been tackled by this method in supercoiled DNA (Bauer and Vinograd 1970, Miyazawa 1985, Sen and Majumdar 1988, Sen and Lahiri 1990). What is more important, for real systems, i.e., DNA with arbitrary base sequences and compositions, the combinatorial method cannot be applied.

### 2. Sequence generating techniques

Sequence generating techniques such as the matrix method and the sequence generating function method avoid cumbersome combinatorial expressions and thus

represent convenient methods for calculating partition functions for complicated systems (Poland and Scheraga 1970, Poland 1978).

The partition function calculated by the matrix method is obtained as

$$Z = e W^m e^+ \quad (2)$$

where  $e$  and  $e^+$  are end vectors, and  $W$  is a matrix of statistical weights. The size of the matrix as well as the index  $m$  depends on the maximum number of correlated units of which the system is composed. Since in supercoiled DNA the condition of invariance of linking introduces coupling between the twisting and writhing deformations throughout the length of the molecule, the conformational states of all the monomer units become in principle correlated. Thus the minimum dimension of the matrix becomes equal to the chain size. For a large molecule such an algorithm offers little advantage.

In sequence generating function method, which in principle is applicable only for chains of infinite length, one directly arrives at a secular equation by neglecting end effects. For example, for the helix-coil transition in homopolynucleotides

$$Z = x_1^N \quad (3)$$

where  $N$  is the number of base pairs in the DNA molecule and  $x_1$  is the largest root of the secular equation

$$h(x)c(x) = 1 \quad (4)$$

where  $h(x)$  and  $c(x)$  are appropriate sequence generating functions for the helix and coil states respectively (Lifson 1964, Majumdar and Thakur 1985). A fundamental assumption required for the derivation of relation (4) is that the statistical weight of a microscopic state is the product of the statistical weights of successive sequences of conformational states in the molecule with each term being dependent on its sequence length and independent of the length of other sequences in the rest of the molecule. Since the supercoiling energy arises from the elastic deformations of the DNA molecule, for small amounts of supercoiling the energy usually shows a harmonic character (Depew and Wang 1975). It is easy to check that with a quadratic dependence of the supercoiling energy on the total deformation the assumption stated above does not hold. It can also be ascertained that the above assumption would hold if the same variation exhibited a linear dependence, or in other words, the rate of change of supercoiling energy with deformation became constant.

From the foregoing it can be concluded that for harmonic dependence of supercoiling energy on the total deformation the sequence generating methods are either inapplicable or inadventagous for the calculation of chain partition functions. But there is one wellstudied case where the supercoiling energy has been observed to deviate from its harmonic nature and to exhibit a linear character.

### 3. Melting of supercoiled DNA

The variation of supercoiling energy  $G_s$  with melting shows a very remarkable nature when a closed circular DNA becomes partially melted. Expression for such a situation has been recently obtained (Sen and Majumdar 1988).

$$G_s = R(\sigma + \theta)^2 / [1 + (\alpha - 1)\theta] \quad (5)$$

where  $R$  is a constant involving the structural parameters of the DNA molecule,  $\sigma$  is the intrinsic supercoil density,  $\theta$  is the fraction of base pairs melted, and  $\alpha$  is the ratio of the rigidity constants of the helix and the coil duplexes. Usually, it is assumed that  $\alpha \gg 1$ .

(i) For  $\sigma \approx 0$  and  $(\alpha - 1)\theta \gg 1$ ,  $G_s$  can be written as

$$G_s \approx R(2\sigma + \theta) / (\alpha - 1) \quad (6)$$

or, in other words, for small values of intrinsic supercoil density the supercoiling free energy tends to a linear form as  $\theta$  increases.

(ii) For  $(\alpha - 1)\theta \ll 1$ ,  $G_s$  can be expressed in the following manner

$$G_s = R(\sigma + \theta)^2 [1 - (\alpha - 1)\theta + (\alpha - 1)^2 \theta^2 - \dots] \quad (7)$$

One can make a complete expansion of  $G_s$  in terms of  $\theta$  from the above expression. We assume that for small values of  $\theta$ , terms involving greater powers of  $\theta$  than the second are negligible. It is interesting to study also the possibility of the coefficient of the quadratic term going to zero leaving only a linear dependence of  $G_s$  on  $\theta$ . Such a calculation yields the relation

$$\sigma = 1 / (\alpha - 1) \quad (8)$$

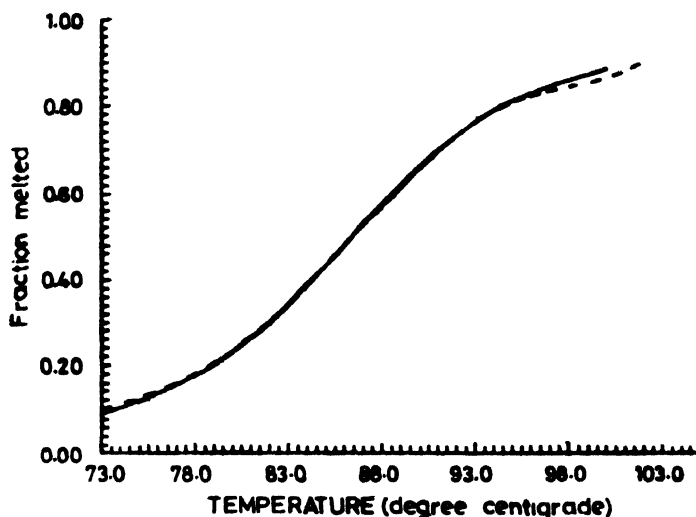
Substitution of the above value of  $\sigma$  into the expression for supercoiling free energy shows that the form of the energy is linear for the entire range of  $\theta$ . The value of  $\alpha$  is usually taken to be about 23 (Sen and Majumdar 1988). This gives a positive value of  $\sigma$  ( $=0.045$ ) at which sequence generating techniques are most suitable for melting studies. However, numerical calculations reveal that this linearity property persists for  $\sigma$  values over a broad range about the value obtained in relation (8). Consequently, for such a range of  $\sigma$  values sequence generating techniques should be applicable to closed circular DNA as well.

In Figure 1, we demonstrate the applicability of the sequence generating technique in describing the melting transition of supercoiled DNA. We compare the melting profile calculated by the combinatorial method of the supercoiled DNA with  $\sigma=0.06$  for which the derivative of the supercoiling energy is almost constant (Sen and Majumdar 1988) with the profile calculated for the same DNA by the sequence generating function method. In the calculation of the secular equation by the sequence generating function method we have assumed that during melting

of the supercoiled DNA large loops are not formed so that loop energy terms do not contribute. For such a situation the secular eq. (4) takes the form

$$(x - s)(x - 1) = \sigma_n s \quad (9)$$

where  $\sigma_n$  is the nucleation parameter or the cooperativity factor coming mainly from the stacking energy term and  $s$  is the equilibrium parameter for the melting of a single base pair which follows another melted base pair. The latter term contains



**Figure 1.** Comparison of the melting profiles of supercoiled DNA ( $\sigma = 0.06$ ) calculated by the sequence generating function method (solid curve) and the combinatorial method (broken curve).

apart from the free energy change in the melting of a single base pair an additional term coming from the change in the supercoiling energy with melting. It can be seen that over almost the entire range of temperature shown the two profiles exhibit excellent agreement.

The fact that  $\partial G_s / \partial \theta$  remains constant over a large range of  $\theta$  is also useful for obtaining the sequence conditional probabilities for a partially melted supercoiled DNA. The sequence conditional probabilities are instrumental for constructing algorithms which calculate the melting probabilities of base pairs in DNA having base heterogeneity (Poland 1978). The above analysis, therefore, helps us to determine the range of applicability of these algorithms in a quite rigorous manner. At present we are investigating by the above outlined method the melting probabilities of specific sequences in supercoiled DNA and plasmids.

### Acknowledgments

The author is grateful to Sree Srikanta Sen, Dr Ashoke Ranjan Thakur and Prof Rabi Majumdar for their keen interest in the problem.

**References**

- Bauer W and Vinograd J 1970 *J. Mol. Biol.* **47** 419  
Depew R E and Wang J C 1975 *Proc. Natl. Acad. Sci.* **72** 4275  
Lifson S 1964 *J. Chem. Phys.* **40** 3705  
Majumdar R and Thakur A R 1985 *Nucl. Acids Res.* **13** 5883  
Miyazawa S 1985 *J. Chem. Phys.* **83** 859  
Poland D 1978 *Cooperative Equilibria in Physical Biochemistry* (Oxford : Clarendon)  
Poland D and Scheraga H A 1970 *Theory of Helix-Coil Transitions in Biopolymers* (New York and London : Academic)  
Sen S and Lahiri A 1990 *Biophys. Chem.* **40** 33  
Sen S and Majumdar R 1988 *Biopolymers* **27** 1479