Eur Biophys J (1996) 25: 9–18

ARTICLE

Y. Z. Chen · E. W. Prohofsky

Sequence and temperature effect on hydrogen bond disruption in DNA determined by a statistical analysis

Received: 20 February 1996 / Accepted: 2 May 1996

Abstract We use the modified self-consistent phonon approximation theory to calculate temperature dependent interbase hydrogen bond disruption profiles for a number of six base pair repeating sequence infinite B-DNA polymers with various guanine-cytosine/adenine-thymine ratios. For comparison we also include results we have obtained in our earlier work on several B-DNA homopolymers, copolymers and a four-base-pair repeating sequence polymer. Our theory gives a statistical estimate of thermal fluctuational disruption probability of individual hydrogen bonds in individual base pairs in DNA as a function of temperature. The calculated probabilities show no sequence dependence at premelting temperatures, in agreement with proton exchange measurements. These probabilities however become very sensitive to base sequence at temperatures close to the observed melting temperatures. Multiphasic critical transitions are found in which a portion of base pairs are disrupted at temperatures below the final disruption temperature. These transitions include localized as well as non-localized base pair opening. The localized transitions involve disruption of a few base-pairs at every other location without large scale base unstacking, and they may not appear in the observed UV curves with current resolution. On the other hand the overall disruption behavior is consistent with observations. The midpoint transition temperatures are close to the observed melting temperatures and these temperatures show the observed linear dependence on guanine-cytosine content. Our calculations indicate that our theory can be used effectively to calculate H-bond disruption behavior of different DNA sequences.

Key words Base pair opening · Chemical bond disruption · Critical transition · DNA melting · Multiphasic transition · Premelting disruption

Y. Z. Chen (✉) · E. W. Prohofsky
Department of Physics, Purdue University, West Lafayette, IN 47907-1396, USA
(Fax 317-494-0706, e-mail yzchen@physics.purdue.edu)

Introduction

Disruption of the interbase hydrogen bonds (H-bonds) in DNA is essential in many biological events related to DNA. An understanding of the effect of DNA sequence and temperature on H-bond disruption is therefore important in the study of the underlying mechanism of these biological processes. Proton exchange experiments indicated that the thermally induced disruption of individual H-bonds at premelting temperatures is insensitive to the sequence of the DNA (Gueron et al. 1987, 1990). In comparison strong sequence dependence is found in the melting of DNA (Lyubchenko et al. 1978; Wada et al. 1980; Wartel and Benight 1985). The denaturation of a random sequence DNA appears to be multiphasic. The regions rich in adenine-thymine (AT) base pairs are found to melt first and the regions rich in guanine-cytosine (GC) base pairs melt last. DNA melting involves massive disruption of H-bonds and unstacking of the bases. This melting is monitored from the change in UV absorption which arises from the unstacking of the bases. While these experiments give a good indication of sequence effects on massive H-bond disruption at melting, the measured profile is indirectly related to the H-bond behavior. In addition the limited resolution does not allow the detection of small changes in base stacking and hinders the ability of the experiment to find localized disruption. It is the purpose of the present work to carry out computer calculations to determine the H-bond disruption profiles of different DNA sequences at melting as well as premelting temperatures.

The observed sequence dependent step-wise denaturation has been successfully predicted by the nearest-neighbor helix-coil transition theory which is based on the Ising two-state model with specific parameters that introduce sequence-specific effects into the melting of DNA (Crothers 1971; Fixman and Freire 1977; Lyubchenko et al. 1978; Wartel and Benight 1985). The parameters are fitted to observed melting behavior in some systems and then reproduce the complex melting profiles in others. In addition to the dependence on the fitted parameters, the helix-coil tran-

sition theory gives no atomic-level information on the melting process. In order to probe the H-bond disruption profile it is necessary to use an atomic level microscopic theory based on atom-atom interactions. One might think that molecular dynamcis (MD) simulation could be used for such a purpose. However its practical use is limited because simulations with current computer capacity can only run for a few nanoseconds for a relatively short DNA sequence with a dozen base pairs. The observed base pair lifetime at room temperature is in the millisecond range (Gueron et al. 1987, 1990). This lifetime is expected to be in the range of microseconds at temperatures close to melting. Therefore MD simulation time is far too short to be useful in the study of the dynamics of base pair dissociation.

We have recently developed a microscopic lattice dynamics – statistical mechanics approach, the cooperative modified self-consistent phonon approximation (MSPA) theory, to calculate the temperature profile of the disruption of interbase H-bonds in DNA polymers (Prohofsky 1995; Chen and Prohofsky 1993, 1994 a). The calculated H-bond critical transition temperatures, at which large scale H-bond disruption occurs, were found to correlate very well with observed melting temperatures for various DNA sequences. In our theory the fluctuational internal motions of a complex anharmonic system such as DNA is approximated by a self-consistently determined effective phonon system (Werthamer 1970) according to the Bogoliubov variational theorem (Callen 1985). Our theory concentrates on the collective excitations around a mean equilibrium conformation and thus also avoids the difficult ground state calculation and problems associated with trapping in local potential minima. The theory is valid up to the onset of massive interbase H-bond disruption which correlates with the melting of the DNA polymer.

Many studies indicate that the melting of infinite repeating sequence DNA polymer is a second order phase transition (Poland and Scheraga 1966; Fisher 1966; Azbel 1974; Chen and Prohofsky 1994 b). As a result the transition involving massive H-bond disruption in these DNAs can be studied from calculations of dissociation from the intact phase. We have applied the MSPA theory to calculate both individual interbase H-bond disruption probability and base pair opening probability (Chen et al. 1991) which are compared to observed amino proton and imino proton exchange (Gueron et al. 1987, 1990). The MSPA theory has also been used to study the effect of the minor groove spine of hydration on the proton exchange behavior of B-DNA Poly(dA) · Poly(dT) (Chen and Prohofsky 1992). By the introduction of a near-neighbor cooperative element the theory successfully predicted the sequence dependent H-bond disruption critical transition consistent with observed melting temperatures of B-DNA homopolymers and copolymers (Chen and Prohofsky 1993, 1994 a).

Unlike random sequence DNA molecules, DNA homopolymers and copolymers do not exhibit multiphasic melting. Although our theory has been successful in predicting sequence dependent H-bond disruption of these homopolymers and copolymers, the theory has not been tested on more irregular DNA sequences. In the present study we extend our calculation to determine the temperature profile of H-bond disruption in a number of six-base-pair repeating sequence B-DNA polymers. We show that our calculation predicts sequence dependent behavior in close correlation with observed premelting and melting behavior of random sequence DNAs. In addition we find more detailed behavior, involving localized disruptions, than available in the observed melting curves with current resolution. Our present study demonstrates that the cooperative MSPA theory as a microscopic lattice dynamics – statistical mechanical theory can be used to probe the onset of massive chemical bond disruption as well as premelting dynamics in biological macromolecules.

Formulation of the long-time dynamics of thermal fluctuational H-bond disruption

The problems investigated combine large systems with long time scale dynamics. This combined burden requires the development of more efficient methods. The principal difference, between the MSPA calculation and a molecular dynamics (MD) calculation, that leads to its greater efficiency in long time scale phenomena is that in MSPA we calculate stationary time independent solutions. This is the same as in simple harmonic normal mode calculations where one also calculates stationary eigenvalues and eigenvectors. In a MD simulation one calculates a continuing set of position and velocity parameters that, in principal, are different at all points in time and that only incidentally may repeat. It is this open endedness in time that expands the MD calculation for phenomena that occur on long time scales. The stationary MSPA calculation is good for all times, i. e. the calculation is the same size independent of the time scale explored. MSPA differs from simple harmonic calculations in the way it handles nonlinearities. For some of the interactions, H-bond and non-bonded, simple harmonic interactions are replaced by effective force constants that can change with the level of excitation. It thus mimics a calculation where higher order nonlinear terms have been incorporated or renormalized into the effective first order term. It does this by self consistent rather than renormalization calculation. The eigenvectors and eigenvalues along with a model of the excitations of each mode then give a mean value for square displacement for each degree of freedom in the original problem. The MD simulation, on the other hand, gives a single trajectory that is open ended in time and that will only give mean values after averaging over many trajectories. Over long time scales mean values contain the more useful information as many trajectories are possible and the specific trajectory chosen is undetermined. Over short time scales individual trajectories may be significant but only if one knows the starting microstate of the system. Such specific trajectory information isn't available in MSPA.

MSPA also has certain advantages in efficiency over Monte Carlo (MC) calculations. In MC calculations one

must sample a large number of points in the phase space of the large system to build up mean values of the system. MSPA also samples phase space in the determination of the effective force constants, but it does so using a particular density matrix that allows closed form integration instead of a more open summation. The density matrix is that of the effective harmonic system and its spatial part is a simple Gaussian distribution in displacement from an equilibrium distance. The Gaussian is characterized by only two parameters, the location of the centroid and the mean square fluctuation which determines the width of the Gaussian. It is only these two parameters that have to be determined self-consistently and given these the integration to determine mean values is very efficient. The self-consistent centroid determination allows for thermal expansion of the distance between particular atoms and both the location of the centroid and the mean fluctuation are amplitude dependent, in contrast to the case of simple harmonic calculations. MSPA uses a variational procedure to determine the best solution which is quasi-harmonic, i. e. a best fit spatial density matrix that is basically Gaussian in the displacements of all pairs of atoms. The point then to be tested is how good is such a form of spatial density matrix for the problems investigated. To the extent that the system studied is quasi-harmonic, MSPA is a very efficient method of theoretical investigation that works for long time scales.

Because MSPA uses an effective force constant it also has similarities to Brownian Motion Dynamics (BMD) calculations. In BMD an effective potential is determined based on energy minimization of other degrees of freedom for a specific trajectory of the degree of freedom being investigated. True minimization is only an appropriate approximation at $T = 0$. MSPA incorporates the statistical fluctuations of all degrees of freedom, including the one being investigated, in the determination of the effective force constant. It also doesn't preselect a particular mode or trajectory for investigation. The eigenvalues and eigenvectors of an entire set of degrees of freedom are treated on an equal footing and self-consistent special degrees of freedom appear rather than being selected beforehand.

The question remains as to whether macromolecular systems are well approximated by a quasi-harmonic theory. The answer to this is best determined by how well the approach, and in particular, how well a Gaussian spatial density matrix predicts experimental observations. The most obvious agreement is that Raman and infrared absorption measurements on most macromolecules do show resonant rather than relaxational modes. That is the most direct evidence that the systems are quasi-harmonic. This is true in DNA for the H-bond breathing modes that seem to be clustered in a band near 85 cm^{-1}. Furthermore MSPA calculations do reproduce these frequencies and their temperature dependence for those modes that have been thoroughly studied. A more stringent test of the method is not whether the frequencies are in agreement but whether the resulting Gaussian density matrix is a good representation of the fluctuations of the system. The mean values of the square fluctuations, like the mode frequencies, are easy to fit with refined force constants. A more demanding test that gets to the heart of the use of a Gaussian is whether the predicted Gaussian tails of the effective density matrix give the correct probability of large displacement. The probability of bond disruption (to be discussed below) is a sensitive measure of the size of this large displacement probability. The agreement of the open bond probability with observation both in the premelting region and in a cooperative version of the theory in predicting melting temperatures for DNA indicates that the tails (the most variable part of the Gaussian) of the systems studied are quite quasi-harmonic. The indications are that MSPA allows calculation of a very good approximation to the spatial part of the density matrix of macromolecules. This in turn allows the calculation of the mean value of a great many variables of the systems.

The principal drawback of MSPA compared to MD or molecular mechanics (MM) calculations is that it requires an initial conformation for the system to be studied. There are associated methods that are mostly undeveloped as yet that can explore changes in conformation but they are most likely less efficient than MD or MM calculations. The efficient use of MSPA is in long time dynamics of systems whose conformation is known. The lack of specific trajectory information is not a problem as the stochastic nature of dynamics of molecules makes this information meaningless for long time scales. Mean behavior is the only useful information for large systems over long times. The calculations to date have not involved very complex models of the H-bond interactions. In particular the Morse potentials used are spherically symmetric potentials and only the distance between end atoms of the H-bond is of importance. This model works very well but the approach can easily be modified to include angular dependence of the H-bond interaction.

For calculations on DNA the actual particle interactions are similar to those used in simulation except that we prefer to use Morse potentials for H-bonded interactions. This gives a simpler intuitive insight into the parameters in the problem as the Morse potential has three parameters that separately can be thought of as the depth of the potential, the width of the potential, and the location of the potential minimum. All these parameters are fit for each H-bond based on three experimental observations: the ionization energy of the bond, the effective force constant as found from refining Raman and infrared data, and the x-ray determined inter atom distances. All of these determinations are made at room temperature and therefore are not fit to any data on the bond disruption or melting behavior. The calculation of the H-bond critical transition temperature at the observed melting temperature does not then arise because of fitted parameters in the way that it does in helix-coil transition theory but does represent a result arising from simple interatom potentials derived from a source independent of bond disruption. The effective force constants used in the theory are then determined from these interatom potentials. The variation of these effective force constants are self-consistently determined such that the free energy of the effective system approaches that of the original system. Therefore the correlation between the ef-

fective and the true system is on the thermodynamic level. A statistical mechanical approach can then be used to determine thermal fluctuational chemical bond disruption probability as well as interatomic vibrational motion at temperatures from the premelting region to the onset of massive interbase H-bond disruption or the melting of the system.

Calculation method

According to the Bogoliubov variational theorem the free energy of an anharmonic system, such as a DNA polymer, described by a Hamiltonian H can be approximated by solution of an effective system described by a Hamiltonian H_{eff} (Callen 1985). The free energy F of the original system is related to the free energy F_{eff} of the effective system by the following inequality:

$$F \le F_{eff} + \langle H - H_{eff} \rangle_0 \tag{1}$$

where $\langle \ldots \rangle_0$ denotes an average taken in the ensemble defined by H_{eff}. By choosing an appropriate solvable system H_{eff} and then minimizing the right band side of the inequality through parameter adjustments, one can obtain the best approximation of the desired free energy F.

Both the free energies and Hamiltonians have two components, one static and one dynamic. The dynamic component is the internal thermal fluctuational vibrational energies and it can be approximated by a phonon Hamiltonian H_0 as described below. The static component is the total potential at equilibrium positions. Therefore $H_{eff} = H_0 + V(r_{eq})$.

Many experiments have shown the vibrational nature of the internal collective excitations and chemical bond motions in DNA (Urabe and Tominaga 1981; Urabe et al. 1985; Powell et al. 1987; Weidlich and Lindsay 1988; Weidlich et al. 1990; Edwards and Liu 1991). In particular vibrational modes associated with base pair disruption melting have been observed (Urabe and Tominaga 1981; Urabe et al. 1985) and calculated (Zhuang et al. 1992; Chen and Prohofsky 1995). It is therefore reasonable to choose a phonon Hamiltonian as the dynamic component for the effective system for DNA and use the lattice dynamics methodology to study the dynamics of bond disruption. This particular Bogoliubov approximation is valid up to the onset of massive bond disruption and it is strongly supported by the good agreement between calculated and observed premelting bond disruption probabilities and the melting temperatures for a number of DNA polymers as shown in our earlier studies (Chen and Prohofsky 1993, 1994 a) and in this work.

The dynamic component of the effective Hamiltonian of a repeating sequence is given by the following lattice dynamics phonon Hamiltonian:

$$H_0 = \sum_{n,i} \frac{1}{2} M_i \dot{\mathbf{u}}_{ni}^2 + \sum_k \sum_{n,m} \sum_j \frac{1}{2} \phi_{n-m,j,k}\, s_{n-m,j,k}^2 \tag{2}$$

where n and m are the indices of unit cells (a unit cell contains a single repeating sequence), i is the index of the atoms in a unit cell, j is the index for internal coordinates and k specifies the types of harmonic motion (valence bond stretch, angle bending, torsion, out of plane bending, H-bonding and motions induced by the nonbonded atom-atom van der Waals and Coulomb interactions). $\mathbf{u}_{ni}$ is the displacement of the ith atom in the nth unit cell in Cartesian coordinates, $s_{n-m,j,k}$ is the internal displacement coordinate and $\phi_{n-m,j,k}$ is the effective internal force constant for a particular bond motion. The force constants include a standard valence force field, hydrogen bond force constants and a nonbonded force field including van der Waals and Coulomb interactions. The valence and long range nonbonded force constants are dependent on the separation of units and these force constants at room temperature are fitted to observed spectra and acoustic modes. The detailed description of the force fields for the atoms in DNA can be found in our earlier publications (Chen and Prohofsky 1995).

In the current work we study idealized repeating sequence B-DNA systems with infinite length and perfect helical symmetry. The helical symmetry inherent in the system can be used to reduce the spatial dimension in the calculation (Chen and Prohofsky 1995). The reduction in spatial dimension introduces a parameter θ, the phase difference between adjacent unit cells, into the equations of motion. The equation of motion in matrix form is given by:

$$\left(\sum_k B_k^+(\theta)\, \Phi_k(\theta)\, B_k(\theta) - \omega_\lambda(\theta)^2\, I \right) q_\lambda(\theta) = 0 \tag{3}$$

where $\Phi_k(\theta)$ is the force constant matrix, $B_k(\theta)$ is the transformation matrix (B-matrix) that transforms the Cartesian displacement coordinates into internal displacement coordinates, and $\omega_\lambda(\theta)$ and $q_\lambda(\theta)$ are the eigenfrequency and eigenvector matrix for the λth normal mode band respectively. From the solution of this equation one can calculate mean vibrational amplitude and bond length for each individual bond as well as other thermodynamic observables.

The force constant of interbase H-bonds, and of other bonds that change with temperature etc., can be determined self-consistently by minimization of the right hand side of the inequality given in Eqn. [1]:

$$\phi = (1-P)C \int_{r_{\min}}^{\infty} dr \exp\{(r - \langle r \rangle)^2 / 2\langle s^2 \rangle\} \frac{d^2 V(r)}{dr^2} \tag{4}$$

where $C^{-1} = \int_{r_{\min}}^{\infty} dr \exp\{-(r - \langle r \rangle)^2 / 2 \langle s^2 \rangle\}$, $\langle s^2 \rangle$ is the mean bond vibrational square amplitude, $\langle r \rangle$ is the mean bond length, $V(r)$ is the potential, $r_{\min}$ is the hardcore innerbound for the potential, P is the bond disruption probability and the factor $1-P$ is introduced to incorporate the cooperative effect of disrupted bonds and transform the theory into a mean-field theory capable of showing cooperative transitions.

According to the Bogoliubov theorem the disruption probability of a chemical bond can be determined from the probability density of the effective system $\rho(r) = \exp\{-(r - \langle r \rangle)^2 / 2 \langle s^2 \rangle\}$ as:

$$P = C \int_{L^{\max}}^{\infty} \rho(r)\, dr \tag{5}$$

where $L^{\max}$ is the maximum stretch length before disruption which is determined from the analysis of the mean bond energy of the effective system (Chen et al. 1991). We have found that the observed DNA base pair opening probability P^{op} correlates very well with the calculated probability for the disruption of all interbase H-bonds in a base pair. Individual H-bond disruption can be regarded as an independent event. Therefore P^{op} can be given as the product of the *Ps* for all the interbase H-bonds in the base pair:

$$P^{op} = \prod_{interbase\ H-bonds} P \tag{6}$$

The nearest neighbor cooperativity is introduced in our theory by scaling the cross-strand base stacking (van der Waals) force constants ϕ^s and the equilibrium length of interbase H-bonds $\langle r \rangle$ in the following way (Chen and Prohofsky 1994 a):

$$\phi^s \rightarrow \left(1 - \sqrt{P_l^{op} P_{l'}^{op}}\right)\phi^s \tag{1}$$

$$\langle r \rangle \rightarrow (1 - P^{uc})\langle r \rangle + P^{uc}\langle r^{op} \rangle \tag{7}$$

where l and l' are index of the base pair and its neighbor respectively, $P^{uc} = (P_{l-1}^{op} P_l^{op} P_{l+1}^{op})^{1/3}$ is a base pair unconstrained probability and $\langle r^{op} \rangle = L^{max} + 2P\sqrt{2\langle s^2 \rangle}$ is the average open bond atom-atom separation.

The effective system formulated is valid up to the onset of massive interbase H-bond disruption. At the point of massive disruption $\phi \rightarrow 0$ and $P^{op} \rightarrow 1$. For systems with mono-step disruption, the onset temperature of massive H-bond disruption T_m is defined as the temperature at which $P^{op} \approx 1/2$ in conformity with the convention of DNA melting. For systems with multi-step disruption the onset temperature of massive H-bond disruption T_m is defined as the midpoint between first significant disruption and final disruption.

The calculation is carried out such that, given the initial force constants and the microscopic structure, the equation of motion of the effective system is solved. The solutions of this equation can then be used to calculate thermodynamic observables and new force constants using the formulas outlined above. The newly calculated force constants are then used as input to start another round of calculation. This iteration process continues until self-consistency, i. e. all the input force constants match output force constants. The solution at self-consistency corresponds to free energy minimization. The calculation was performed on an IBM RS6000 570 server. The equations of motion of the effective system are solved by the use of the IMSL Hermitian complex diagonalizations subroutine DEVCHF.

Results and discussion

In this work we consider five six-base-pair repeating sequence B-DNA polymers with different GC contents. These 6-polymers are: Poly d(GTATAT) · Poly d(ATATAC), Poly d(GTATAC) · Poly d(GTATAC), Poly d(GTATGC) · Poly d(GCATAC), Poly d(GCTAGC) · Poly d(GCTAGC) and Poly d(GCGAGC) · Poly d(GCTCGC). For comparison we also include results for several B-DNA homopolymers and copolymers (Chen and Prohofsky 1994 a) and a B-DNA 4-polymer (Chen and Prohofsky 1996) that we have studied in earlier work. A well organized spine of hydration has been found to exist in the minor groove of the B-DNA homopolymer Poly d(A) · Poly d(T) (Chuprina et al. 1991). We therefore included this spine of hydration in our calculation for this polymer. Our earlier study (Chen and Prohofsky 1992 b, 1993) indicates that this hydration spine is most likely responsible for the observed anomalous premelting and melting behavior of this homopolymer. The coordinates of all these DNA polymers are generated from the standard B-DNA structure obtained from fiber X-ray diffraction studies (Chandrasekaran and Arnott 1989). Our calculation corresponds to a nominal salt concentration of ~0.05 M NaCl (Chen and Prohofsky 1992 b).

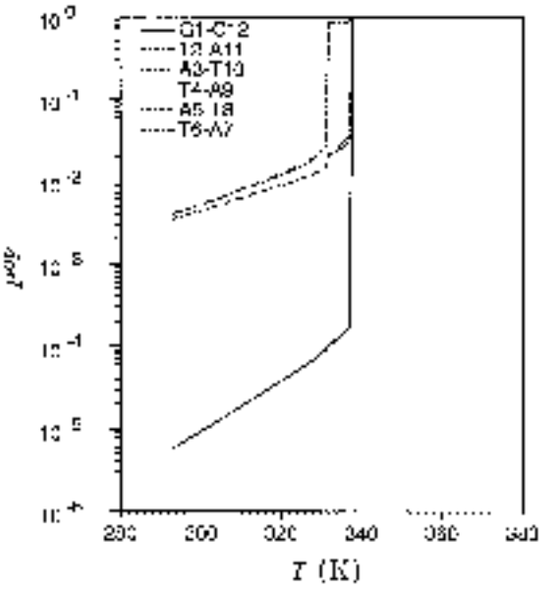

Fig. 1 Base pair opening probability P^{op} as a function of temperature for the base pairs in Poly d(GTATAT) · Poly d(ATATAC). The nomenclatures for the bases are G1-T2-A3-T4-A5-T6 on one strand and A7-T8-A9-T10-A11-C12 on the opposite strand. The *solid line* is for the G1-C12 pair, the chain-dotted lines for T2-A11 and T6-A7 pairs, the *dashed lines* for A3-T10 and A5-T8 pairs and the *dotted line* for T4-A9 pair

Sequence effect on premelting base pair opening probability

Our calculated base pair opening probability (P^{op}) as a function of temperature for the five 6-polymers are shown in Fig. 1 through Fig. 5 respectively. The P^{op} for the homopolymers, copolymers and the 4-polymer obtained in our earlier work are shown in Fig. 6 and Fig. 7. Figure 6

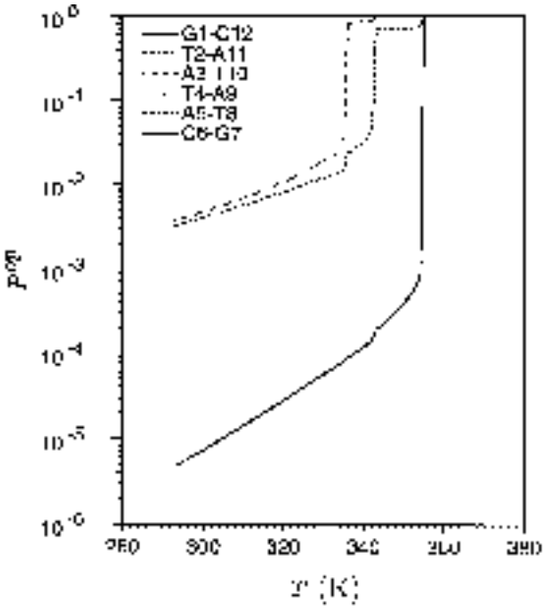

Fig. 2 Base pair opening probability P^{op} as a function of temperature for the base pairs in Poly d(GTATAC) · Poly d(GTATAC). The nomenclatures for the bases are G1-T2-A3-T4-A5-C6 on one strand and G7-T8-A9-T10-A11-C12 on the opposite strand. The *solid lines* are for G1-C12 and C6-G7 pairs, the *dashed lines* for T2-A11 and A5-T8 pairs and the *dotted lines* for A3-T10 and T4-A9 pairs

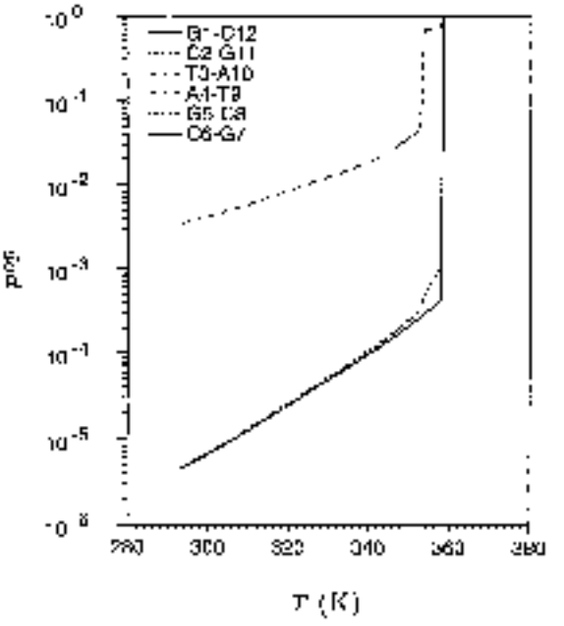

Fig. 4 Base pair opening probability P^{op} as a function of temperature for the base pairs in Poly d(GCTAGC) · Poly d(GCTAGC). The nomenclatures for the bases are G1-C2-T3-A4-G5-C6 on one strand and G7-C8-T9-A10-G11-C12 on the opposite strand. The *solid lines* are for the G1-C12 and C6-G7 pairs, the *dashed lines* for C2-G11 and G5-C8 pairs and the dotted lines for T3-A10 and A4-T9 pairs

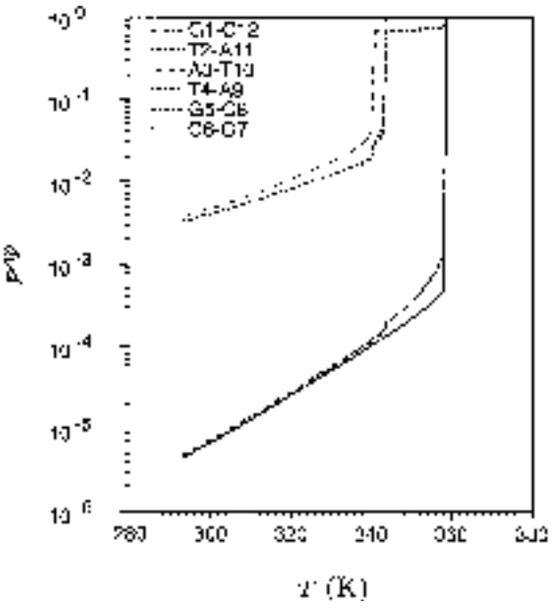

Fig. 3 Base pair opening probability P^{op} as a function of temperature for the base pairs in Poly d(GTATGC) · Poly d(GCATAC). The nomenclatures for the bases are G1-T2-A3-T4-G5-C6 on one strand and G7-C8-A9-T10-A11-C12 on the opposite strand. The *solid line* is for the C6-G7 pair, the chain-dotted lines for G1-C12 and G5-C8 pairs, the dashed lines for T2-A11 and T4-A9 pairs and the dotted line is for the A3-T10 pair

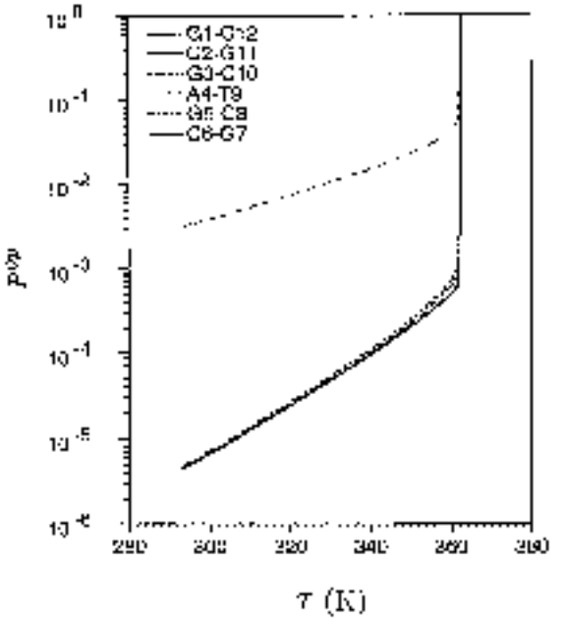

Fig. 5 Base pair opening probability P^{op} as a function of temperature for the base pairs in Poly d(GCGAGC) · Poly d(GCTCGC). The nomenclatures for the bases are G1-C2-G3-A4-G5-C6 on one strand and G7-C8-T9-C10-G11-C12 on the opposite strand. The *solid lines* are for G1-C12, C2-G11 and C6-G7 pairs. The *dashed line* is for G3-C10, the *dotted line* is for A4-T9 and the *chain-dotted line* is for the G5-C8 pair

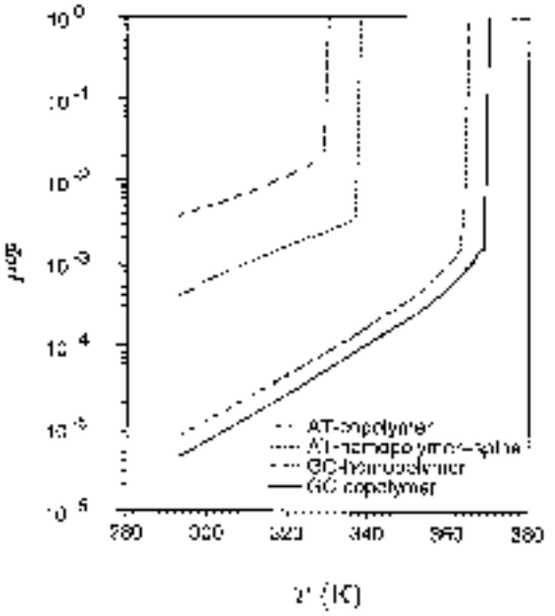

Fig. 6 The base pair opening probability P^{op} as a function of temperature for the base pairs in the B-DNA mono-base-pair polymers studied in our earlier work. The *dotted line* is for the P^{op} of AT pairs in Poly d(AT) · Poly d(AT), the *dashed line* is for the P^{op} of the AT pair in Poly d(A) · Poly d(T) with the hydration spine, the *chain-dotted line* is for the P^{op} of the GC pair in Poly d(G) · Poly d(C), and the *solid line* is for the P^{op} of the GC pairs in Poly d(GC) · Poly d(GC)

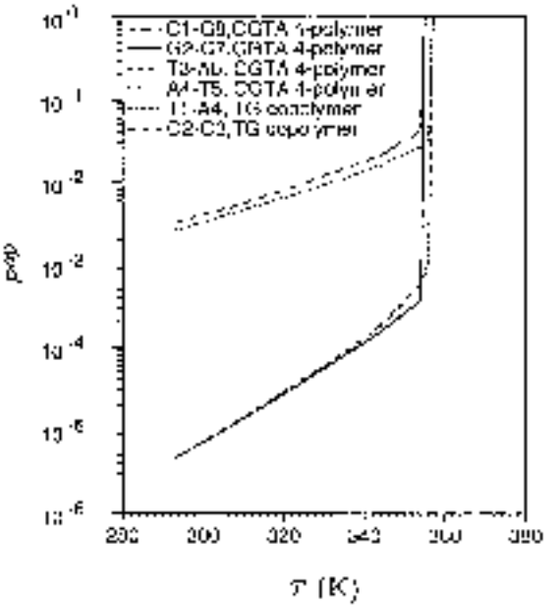

Fig. 7 The base pair opening probability P^{op} as a function of temperature for the base pairs in the B-DNA GC-AT polymers studied in our earlier work. The two solid lines and the *two dotted lines* are for the P^{op} of the respective GC and AT pairs in Poly d(CGTA) · Poly d(TACG). The *dashed line* and *chain-dotted line* is the P^{op} for the respective AT and GC pair in Poly d(TG) · Poly d(CA)

contains P^{op}s for the mono-base pair type polymers and Fig. 7 contains P^{op}s for the polymers with a mixture of AT and GC pairs. From these figures we find that, except for the homopolymer Poly d(A) · Poly d(T), the P^{op} for both the AT and GC pairs in these polymers is insensitive to the sequence at premelting temperatures. At room temperature the calculated P^{op}s for the AT pairs are of the order of ~10^{-3}. These P^{op}s follow roughly a specific logarithm dependence with temperature at temperatures below 320 K. Above that temperature these P^{op} begin to show apparent sequence dependence. The calculated P^{op} for Poly d(A) · Poly d(T) is ~10^{-5} at room temperature. We have shown (Chen and Prohofsky 1992 a, 1993) that the significant difference between the P^{op} in this homopolymer and that of the AT pairs in other polymers arises because of the stabilizing effect from the spine of hydration attached to the minor groove of Poly d(A) · Poly d(T).

The P^{op} for the GC pairs in these DNA polymers also show sequence insensitivity at premelting temperatures. From Fig. 1 to Fig. 7 we find that the calculated P^{op}s for all the GC pairs are ~10^{-6}. These P^{op}s again roughly follow a specific logarithm dependence with increasing temperature at temperatures below 340 K. At temperatures above 340 K these P^{op}s begin to show sequence dependence.

The predicted sequence insensitivity and the calculated value of P^{op}s at premelting temperatures are in good agreement with experiments. Imino proton exchange studies (Gueron et al. 1987, 1990) have shown that the lifetime and opening probability of a base pair is determined mainly by the nature of that base pair. These quantities are not sensitive to the nature of neighboring base pairs. Two different values of P^{op} have been observed for AT (or AU) pairs in diffeent systems, one of the order of 10^{-3} (McGhee and Hippel 1977; Gueron et al. 1987, 1990) and the other of the order of 10^{-5} (Gueron et al. 1987, 1990). Our calculated P^{op} for the AT pairs without the hydration spine is in agreement with the former and our calculated P^{op} for the AT pairs with the spine is close to the latter. We have pointed out in our earlier study (Chen and Prohofsky 1992 a, 1993) that the AT pairs with an observed probability of 10^{-5} are likely to have a spine of hydration in their minor groove. The stabilizing effect of this hydration spine contributes to the lower probability. On the other hand those AT pairs whose observed probability is of the order of 10^{-3} are most likely without the hydration spine attached to their minor groove.

Sequence dependent profiles of H-bond disruption critical transition

While insensitive to the neighboring sequence at premelting temperatures, the calculated base pair opening probabilities are found to be very sensitive to the sequence at temperatures near the observed melting temperatures as shown in Fig. 1 through Fig. 7. We find from Fig. 6 and Fig. 7 that there exists a single critical temperature for each

homopolymer, copolymer and 4-polymer studied at which the P^{op} increases dramatically from below 10^{-1} to 1. This dramatic increase indicates the onset of massive interbase H-bond disruption and the critical temperature correlates with the melting temperature of the polymer. Although the P^{op} of an AT pair in both Poly d(TG) · Poly d(CA) and Poly d(CGTA) · Poly d(TACG) is significantly different from that of the GC pair, the critical transition is monophasic. This is consistent with the observed melting behavior. UV melting measurement on alternating AT-GC sequence B-DNA copolymers showed that the transition is monophasic (Wells et al. 1970). Although we have not found a report on the experimental study of the melting of Poly d(CGTA) · Poly d(TACG), nonetheless we can compare our calculation with observed melting profiles of a number of 3-polymers (Wells et al. 1970). Experiments on the 3-polymers Poly d(TTC) · Poly d(GAA), Poly d(TAC) · Poly d(GTA), Poly d(TTG) · Poly d(GAA) and Poly d(ATC) · Poly d(GAT) showed that the melting of these polymers is monophasic.

From Fig. 1 to Fig. 5 we find that the disruption profiles of the 6-polymers show very interesting features. Except for Poly d(GCGAGC) · Poly d(GCGAGC), the probability curves indicate massive AT pair disruption at certain temperatures substantially below the final GC pair disruption temperature. This is consistent with observed melting behavior in random sequence DNAs which showed that AT rich regions melt first (Lyubchenko et al. 1978; Wada et al. 1980; Wartel and Benight 1985). To better probe the multiphasic behavior of these 6-polymers we plot in Fig. 8 the average base pair opening probability $\langle P^{op} \rangle$ of these polymers along with that of the GC-AT mixing copolymers and the 4-polymer. In Fig. 8 the solid-lines correspond to the $\langle P^{op} \rangle$ of those polymers exhibiting multiphasic disruption behavior and the dotted lines correspond to the $\langle P^{op} \rangle$ of those polymers that do not exhibit multiphasic behavior. One can see from Fig. 8 that the average probability curves for the four 6-polymers exhibit multiphasic transitions reminiscent of the observed UV multiphasic curves.

Several different types of critical disruption transitions are found in the AT-GC mixed polymers shown in Fig. 8. The types of transition include triphasic, biphasic and monophasic. Triphasic transitions are found in both Poly d(GTATAC) · Poly d(GTATAC) and Poly d(GTATGC) · Poly d(GTATGC). As shown in Fig. 2 and Fig. 3 the first phase corresponds to the disruption of the interbase H-bonds in middle AT pair(s) in these polymers, the second phase the disruption of the interbase H-bonds in the AT pairs adjacent the GC pairs, and the final phase the disruption of the interbase H-bonds in the GC pairs. Biphasic transitions are found in both Poly d(GTATAT) · Poly d(ATATAC) and Poly d(GCTAGC) · Poly d(GCTAGC). As shown in Fig. 1 and Fig. 4 the first phase is again related to the disruption of the interbase H-bonds in the middle AT base pairs and the final phase the disruption of interbase H-bonds in the remaining base pairs. The melting transitions in Poly d(GCGAGC) · Poly d(GCGAGC) as well as in the copolymer Poly d(TG) · Poly d(CA) and 4-polymer Poly d(CGTA) · Poly d(TACG) are monophasic.

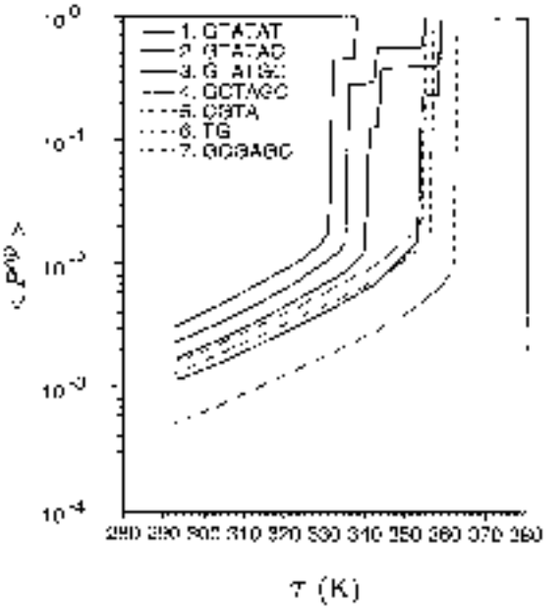

Fig. 8 Average base pair opening probability $\langle P^{op} \rangle$ as a function of temperature for the DNA polymers with AT-GC mixtures studied in the present work and in our earlier work. The *solid lines* are for DNA polymers exhibiting multiphasic melting transition and the *dashed lines* are for those with monophasic transition. The *upper solid line* is for Poly d(GTATAT) · Poly d(ATATAC), the *second solid line* for Poly d(GTATAC) · Poly d(GTATAC), - - - - - the *third solid line* for Poly d(GTATGC) · Poly d(GCATAC) and the lower solid line is for Poly d(GCTAGC) · Poly d(GCTAGC). - - - - - - - The *upper dashed line* is for Poly d(CGTA) · Poly d(TACG), the *middle dashed line* is for Poly d(TG) · Poly d(CA) and the *lower dashed line* is for Poly d(GCGAGC) · Poly d(GCTCGC)

Some of our calculated critical transitions are localized disruptions involving base pairs at every other location. These transitions do not cause appreciable base unstacking and they are difficult to detect in the UV absorbance curves with current resolution. The UV absorbence arises from the π-π^* electronic transition in the bases. An increase in the absorbance reflects a change in base stacking. The limit in resolution hinders the ability of melting experiments to detect localized disruptions that do not cause substantial base unstacking. Therefore our theory predicts more detailed disruption behavior than can be detected from current experimental methods. It would be interesting to see if novel experimental methods can be developed to detect localized transitions near melting and to test our predictions.

Our calculated overall critical transition temperatures T_ms for the polymers studied turn out to correlate fairly well with the observed melting temperatures for DNA polymers with similar GC content. The comparison is shown in Table 1 and the experimental data in Table 1 are from Wells et al. (1970) and from Inman and Baldwin (1964). We have shown in our earlier study (Chen and Prohofsky 1994a) that our calculated T_ms for B-DNA homopolymers and copolymers are in good agreement with observed melting temperatures. Although we have not found

Table 1 Calculated midpoint critical transition temperature (T_m^{theor}) of the B-DNA polymers studied in this work and in our earlier work along with observed melting temperature (T_m^{expt}) for similar B-DNA polymers. All the data correspond to a salt concentration of 0.05 NaCl. I refers to Inman and Baldwin 1962, W refers to Wells et al. 1970

GC content (%)	Sample	Transition type	T_m^{theor} (K)	T_m^{expt} (K)	Ref.
0.0	Poly d(AT) · Poly d(AT)	Monophasic	330	327	W
0.0	Poly d(A) · Poly d(T)+spine	Monophasic	338	335	I
16.7	Poly d(GTATAT) · Poly d(ATATAC)	Biphasic	335		
33.3	Poly d(GTATAC) · Poly d(GTATAC)	Triphasic	345		
33.3	Poly d(TTC) · Poly d(GAA)	Monophasic		344	W
33.3	Poly d(TAC) · Poly d(GTA)	Monophasic		346	W
33.3	Poly d(TTG) · Poly d(GAA)	Monophasic		349	W
33.3	Poly d(ATC) · Poly d(GAT)	Monophasic		350	W
50.0	Poly d(GTATGC) · Poly d(GCATAC)	Triphasic	351		
50.0	Poly d(TC) · Poly d(GA)	Monophasic	354	350	W
50.0	Poly d(CGTA) · Poly d(TACG)	Monophasic	355		
50.0	Poly d(TG) · Poly d(CA)	Monophasic	357	356	W
66.7	Poly d(GCTAGC) · Poly d(GCTAGC)	Biphasic	359		
83.3	Poly d(GCGAGC) · Poly d(GCTCGC)	Monophasic	363		
100.0	Poly d(G) · Poly d(C)	Monophasic	365	367	I
100.0	Poly d(GC) · Poly d(GC)	Monophasic	370	378	W

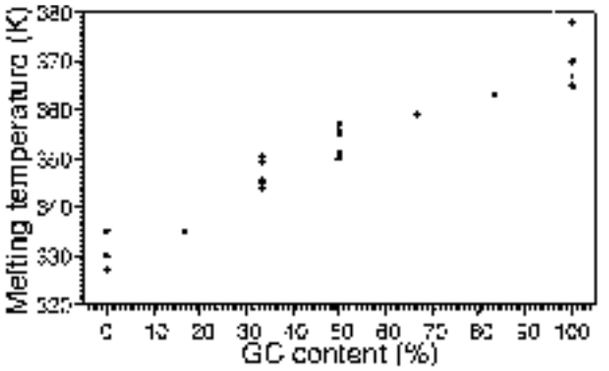

Fig. 9 Calculated midpoint critical transition temperatures (*circles*) and observed melting temperatures (pluses) for the DNA polymers listed in Table 1 as a function of Guanine-cytosine content

an observed melting temperature for the 4-polymer and the 6-polymers listed in Table 2, nonetheless we can compare our calculated T_ms with observed values of DNA polymers with the same GC content. From Table 1 we find that the T_m for Poly d(GTATAC) · Poly d(GTATAC) is 345 K at 0.05 NaCl. The observed T_ms for a number of 3-polymers with the same GC content range from 344 K to 350 K at the same salt concentration (Wells et al. 1970). The calculated T_m for Poly d(GTATGC) · Poly d(GCATAC) is 351 K at 0.05 NaCl. This value is compared to the observed T_m of 350 K for Poly d(TC) · Poly d(GA) and 356 K for Poly d(TG) · Poly d(CA) at the same salt concentration.

We have not found an experimental study for DNA polymers with GC contents similar to Poly d(GTATAT) · Poly d(ATATAC), Poly d(GCTAGC) · Poly d(GCTAGC) and Poly d(GCGAGC) · Poly d(GCTCGC) to compare with. However from Fig. 9 we find that our calculated T_m for these two 6-polymers along with the T_m for other polymers first nicely to a linear relationship with GC content. This linear-relationship has been observed in melting measurements of many DNA samples (Marmur and Doty 1962). This correlation further demonstrates that the near-neighbor cooperative effect in DNA is properly described in our cooperative MSPA theory.

Conclusion

The interbase H-bond disruption profiles of several six-base-pair repeating sequence infinite B-DNA polymers were calculated by the use of a lattice dynamics – statistical mechanics method, the cooperative MSPA approach. The calculated overall sequence-temperature behaviors at both the premelting and melting temperatures are consistent with observations. In addition our theory predicts more detailed critical disruption behavior than available from melting experiments.

It should be emphasized that all the initial parameters of the cooperative MSPA theory are fitted to observed spectra and chemical bond lengths at room temperature. Statistical algorithms are then used to calculate the force constants, the variation of these parameters and the thermal fluctuational chemical bond disruption probabilities at higher temperatures. The agreement with observed T_m is not a result of fitting any parameters to the observed melting data. Since the melting of a long DNA polymer is a classical critical second order phase transition, the disruption probabilities as order parameters can be introduced into the effective interaction potentials. This then gives the sequence dependent near-neighbor cooperative effect. Unlike the nearest-neighbor helix-transition theory no explicit near-neighbor parameters are introduced in our theory to fit the melting.

Because of the complexity of DNA, simulation can not be used to deal with long time dynamics of bond disrup-

tion base pair dissociation. As demonstrated in the present work these dynamics can be efficiently studied by the cooperative MSPA based on the lattice dynamics approach and statistical mechanics algorithm. The fair agreement between our calculated behavior and observations shows the applicability of the cooperative MSPA theory in the study of massive bond disruption as well as premelting dynamics of DNA and other biological macromolecules.

Acknowledgements This work is supported in part by ONR Grant N00014-92-K-1232.

References

Azbel MY (1974) Generalized one-dimensional Ising model for polymer thermodynamics. J Chem Phys 62: 3635–3641

Callen H (1985) Thermodynamics and an introduction to thermostatics, 2nd edn. Wiley and Sons, New York, pp 433–435

Chandrasekaran R, Arnott S (1989) The structures of DNA and RNA helices in oriented fibers. In: Sanger W (ed) Landolt-Bornstein Numerical Data and Functional Relationships in Science and Technology, vol VII/1b. Springer, Berlin Heidelberg New York, pp 31–170

Chen YZ, Zhuang W, Prohofsky EW (1991) Premelting thermal fluctuational interbase hydrogen bond disrupted states of a B-DNA Guanine-Cytosine base pair. Significance for amino and imino proton exchange. Biopolymers 31: 1273–1281

Chen YZ, Prohofsky EW (1992a) The role of a minor groove spine of hydration in stabilizing Poly(dA)-Poly(dT) against fluctuational interbase H-bond disruption in the premelting temperature regime. Nucleic Acids Res 20: 415–419

Chen YZ, Prohofsky EW (1992b) Salt-dependent stability of Poly(dG)-Poly(dC) with potential of mean force Coulomb interactions. Biopolymers 32: 1123–1127

Chen YZ, Prohofsky EW (1993) Differences in melting behavior between homopolymers and copolymers of DNA: role of nonbonded forces for GC and the role of hydration spine and melting transition for AT. Biopolymers 33: 797–812

Chen YZ, Prohofsky EW (1994a) Near-neighbor effects in cooperative modified self-consistent phonon approximation melting in DNA. Phys Rev E 49: 873–881

Chen YZ, Prohofsky EW (1994b) First- or second-order transition in the melting of repeating sequence DNA. Biophys J 66: 202–206

Chen YZ, Prohofsky EW (1995) Sequence and temperature dependence of the interbase hydrogen bond breathing modes in B-DNA polymers. Comparison with low frequency Raman peaks and their role in helix melting. Biopolymers 35: 573–582

Chen YZ, Prohofsky EW (1996) Melting profile and temperature dependent binding constant of an anticancer drug daunomycin-DNA complex. Eur Biophys J 24: 203–212

Chuprina VP, Heinemann U, Nurislamov AA, Zielenkiewicz P, Dickerson RE (1991) Molecular dynamics simulation of the hydration shell of a B-DNA decamer reverts two main types of minor-groove hydration depending on groove width. Proc Natl Acad Sci USA 88: 593–597

Crothers DM (1971) Statistical thermodynamics of nucleic acid melting transitions with coupled binding equilibria. Biopolymers 10: 2147–2160

Edwards G, Liu C (1991) Sequence dependence of low-frequency Raman-active modes in nucleic acids. Phys Rev A 44: 2709–2717

Fisher ME (1966) Effect of the excluded volume on phase transitions in biopolymers. J Chem Phys 45: 1496–1473

Fixman M, Freire JJ (1977) Theory of DNA melting curves. Biopolymers 16: 2693–2704

Gueron M, Kochoyan M, Leroy JL (1987) A single mode of DNA base-pair opening drives imino proton exchange. Nature 328: 89–92

Gueron M, Charretier E, Hagerhorst J, Kochoyan M, Leroy JL, Moraillon A (1990) Applications of imino proton exchange to nuclei acid kinetics and structures. In: Sarma RH, Sarma MH (eds) Structure and Methods, vol 3: DNA and RNA. Adenine Press, New York, pp 113–137

Inman RB, Baldwin RL (1964) Helix-random coil transitions in DNA homopolymer pairs. J Mol Biol 8: 452–469

Lyubchenko YL, Vologodskii AV, Frank-Kamenetskii MD (1978) Direct comparison of theoretical and experimental melting profiles for RF II Phi×174 DNA. Nature 271: 28–31

Marmur J, Doty P (1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. J Mol Biol 5: 109–118

McGhee JD, Hippel PH von (1977) Formaldehyde as a probe of DNA structure. 4. Mechanism of the initial reaction of formaldehyde with DNA. Biochemistry 16: 3276–3293

Poland D, Scheraga HA (1966) Occurrence of a phase transition in nucleic acid models. J Chem Phys 45: 1461–1469

Powell JW, Edwards GS, Genzel L, Kremer F, Wittlin A, Kubasek W, Peticolas W (1987) Investigation of far-infrared vibrational modes in polynucleotides. Phys Rev A 35: 3929–3939

Prohofsky EW (1995) Statistical mechanics and stability of macromolecules. Cambridge Univ Press, Cambridge

Urabe H, Tominaga Y (1981) Low frequency Raman spectra of DNA. J Phys Soc Japan 50: 3543–3544

Urabe H, Hayashi H, Tominaga Y, Nishimura Y, Kubota K, Tsuboi M (1985) Collective vibrational modes in molecular assembly of DNA and its application to biological systems. Low frequency Raman spectroscopy. J Chem Phys 82: 531–535

Wada A, Yabuki S, Husimi Y (1980) Fine structure in the thermal denaturation of DNA: high temperature-resolution spectrophotometric studies. CRC Crit Rev Biochem 9: 87–144

Wartel RW, Benight AS (1985) Thermal denaturation of DNA molecules: A comparison of theory with experiment. Phys Reports 126: 67–107

Weidlich T, Lindsay SM (1988) Raman study of the low-frequency vibrations of polynucleotides. J Phys Chem 92: 6479–6482

Weidlich T, Lindsay SM, Rui Q, Rupprecht A, Peticolas WL, Thomas GA (1990) A Raman study of low frequency intrahelical modes in A-, B-, and C-DNA. J Biomol Struct Dyn 8: 139–171

Wells RD, Larson JE, Grant RC, Shortle BE, Cantor CR (1970) Physicochemical studies on polydeoxyribonucleotides containing defined repeating nucleotide sequences. J Mol Biol 54: 465–497

Werthamer NR (1970) Self-consistent phonon formulation of anharmonic lattice dynamics. Phys Rev B 1: 572–581

Zhuang W, Chen YZ, Prohofsky EW (1992) Description of base motion and role of excitations in the melting of Poly(dG)-Poly(dC). J Biomol Struct Dyn 10: 403–414