Near-neighbor effects in cooperative modified self-consistent phonon approximation melting in DNA

Y. Z. Chen and E. W. Prohofsky
Department of Physics, Purdue University, West Lafayette, Indiana 47907
(Received 2 August 1993)

The modified self-consistent phonon approximation approach is a microscopic theory describing DNA thermal fluctuational motion both in the premelting temperature regime and in the helix-coil transition region. In the present paper we show that the near-neighbor effects can be introduced into the theory by a simple extension in the logic of the cooperativity. The present theory can give base-sequence effects on melting as well as premelting behavior in agreement with experiments. These effects arise from the same interatomic interaction model used in our earlier calculations [Biopolymers **33**, 351 (1993); **33**, 797 (1993)] and the results do not require the explicit introduction of near-neighbor parameters. The work then becomes a no-parameter derivation of near-neighbor effects in DNA.

PACS number(s): 87.15.By, 63.20.Dj, 63.70.+h, 87.10.+e

I. INTRODUCTION

The very successful nearest-neighbor helix-coil transition theory [1–4] incorporates specific parameters that introduce sequence-specific effects into the melting of DNA. These near-neighbor parameters, as well as other parameters in that theory, are fitted to observed melting behavior and then reproduce melting behavior. The modified self-consistent phonon approximation (MSPA) approach, on the other hand, is a microscopic theory where the degrees of freedom are the displacement of atoms rather than the higher-level abstractions of open or closed base pairs [5–8]. The parameters used in MSPA all relate to interatom interactions and all these parameters are fitted to dynamic data, Raman and ir frequencies, and x-ray lengths, at room temperature. Statistical algorithms are then applied to calculate thermal probabilities at higher temperatures, but no parameters are fitted to temperature-dependent observations. This method shows cooperative melting when cooperative links are included but the only fitted inputs are the interatom interactions fitted at room temperatures. In this paper we show that base-sequence effects on melting can come from a simple extension of cooperativity in MSPA calculations and that the near-neighbor effects then arise from the same interaction model used in earlier calculations. The results do not require the explicit introduction of near-neighbor parameters.

The simple extension only involves two additions to the logic of the older method. One is the realization that the stacking interactions between two base pairs are influenced by both of the pairs involved, rather than just one. The second realization is that a given base pair is somewhat constrained in its motion by the pairs on either side of it and that these pairs can be made of different bases. Both the distinctions listed are meaningless in the case where all base pairs are of the same kind, but become important for the DNA polymers studied here where all four bases are included.

In an earlier work [7] we have illustrated that the MSPA theory with cooperative effects associated with open bonds and open base pairs can be used to evaluate the melting temperatures as well as premelting base pair opening probabilities of DNA polymers. In another work [8] we used MSPA to study the melting and premelting behavior of several DNA copolymers and homopolymers with mono-base-pair sequences. We showed that the difference in the melting temperatures between the guanine-cytosine (GC) homopolymer poly(dG)·poly(dC) and GC copolymer poly[d(G-C)]·poly[d(G-C)] is due to the difference in the cross strand nonbonded interactions. We also showed that the observed higher melting temperature of the adenine-thymine (AT) homopolymer poly(dA)·poly(dT) with respect to that of the AT copolymer poly[d(A-T)]·poly[d(A-T)] is caused by the enhanced thermal stability of the homopolymer from a well-defined spine of hydration attached to its minor groove. Our calculated opening probabilities and melting temperatures for these DNA polymers are in fair agreement with experimental observations. That work, however, was limited to mono-base-pair polymers. In the present paper we extend our work to include those copolymers that have alternating AT-GC sequences. We show that a proper consideration of the cooperative effect between base pairs of different types as well as of the same type is essential in obtaining the observed melting behavior as well as premelting behavior for the alternating AT-GC sequences.

The reason that cooperative effects can easily be introduced into MSPA is that universality arguments indicate that cooperative critical behavior at critical transitions depends on gross properties of the system. In MSPA the cooperativity occurs by explicitly incorporating the opening probabilities. This is based on the consideration that the effects associated with the disruption of individual bonds should be included in a mean field theory such as MSPA in order to properly describe the melting behavior in the transition region where bond disruption probabilities are large. The cooperative effect is incorporated into the theory by the use of probability-weighted linear combination of intact state and open state effective force constants and bond lengths. The success of our theory in

predicting both the observed melting temperatures and the observed premelting opening probabilities for both AT and GC mono-base-pair DNA polymers does seem to support this use of the universality argument.

In dealing with heterogeneous sequences one must introduce some sort of nearest-neighbor sequence dependence into the cooperativity in a similar way as that in the helix-coil transition theory. This sequence dependence can be easily introduced in MSPA by incorporating the opening probabilities of nearest-neighbor base pairs as well as the probability of the base pair considered in the MSPA algorithm. This is consistent with the assumption used in the helix-coil transition theory that the stability of a base pair is influenced by the stability of its nearest-neighbor base pairs [1–4]. We find that the use of a geometric mean of these probabilities in the cooperative element meets all the requirements and gives results in agreement with observations. In this paper we specifically consider two *B*-DNA copolymers: poly[d(T-C)]·poly[d(G-A)] and poly[d(T-G)]·poly[d(C-A)]. We calculate the base pair opening probabilities of these two polymers at temperatures in the premelting region and the disruption of the interbase H bonds that precedes the helix-coil transition. The results are then compared with observations. These results are also compared to those of DNA mono-base-pair homopolymers and copolymers we studied earlier. The general agreement between our calculations and observations seems to indicate that the formulation of near-neighbor cooperative effect presented in this paper is effective in describing the true bond disruption behavior of DNA polymers.

As will be shown below, our detailed calculations depend on the screw axis symmetry of the helix and for this reason alone break down at the point of disruption of the helical structure. It is therefore not a theory of the helix-coil transition but only a theory of the cooperative dissociation of the interbase connections in a helical structure. It can only correctly describe the helix-coil transition if this cooperative disruption is the rate limiting event that controls the one-sided helix-coil transition. That this is the case is reasonable but can only be justified by the agreement between the predictions of this model and experimental observations.

II. THEORY

The MSPA theory is a theory that emulates a canonical ensemble system in a self-consistent box [5–7]. This theory is based on the self-consistent phonon formulation of anharmonic lattice dynamics [9]. In MSPA we describe a DNA at the atomic level of detail. For convenience we treat the hydrogen atoms as if they are bound to their parent atoms and their masses are added to these parent atoms. The coordinates of the double helices are from fiber and crystal data and the system is represented by a Hamiltonian similar to the standard form used in simulations [8]. In principle we conceive of a DNA helix with all the interatomic interactions represented by realistic bounded potentials. To calculate statistical behavior these potentials are then replaced, at an atomic level of detail, by MSPA self-consistent unbounded interactions, i.e., connected by effective force constants. These force constants are calculated self-consistently at each temperature.

Some of the force constants change little with temperature and they can be assumed independent of temperature. These include the valence force constants and the long-range nonbonded force constants other than those of the cross strand base stacking force constants. The valence force fields describe bond stretching, angle bending and torsion, etc. The valence force constants for the bases and backbones used in this study are from Refs. [10,11], respectively. The long-range nonbonded force constants are necessary to reproduce the observed acoustic modes [12] and these force constants are formulated in Refs. [12,13]. To further simplify the calculation we assume that the cross strand base stacking force constants have the same temperature dependence as that of the average of the interbase H-bond force constants. It is the construction of that average that is the new element. The self-consistent loop calculation is thus reduced to the interbase H-bond degrees of freedom, although the normal mode calculation is in the full dimensionality of the system.

In dealing with a repeating sequence DNA polymer such as a homopolymer or a copolymer one can utilize the helical symmetry inherent in the system to further reduce the calculation. We therefore divide a DNA sequence into unit cells. A unit cell contains a single repeating section which is composed of one or more base pairs and the associated backbones. For example, the unit cell of a homopolymer contains one base pair and the unit cell of a copolymer contains two base pairs. The normal mode calculation can then be reduced to a number of calculations each of the dimensionality of a single base pair. We define a DNA homopolymer or a copolymer as a sequence with an infinite number of base pairs. The reduction of the calculation into a unit cell then results in the introduction of a phase angle θ in the MSPA formalism. This phase angle resides in the first Brillouin zone: $-\pi < \theta \leq \pi$.

We have shown [5,6] that the force constant of an intact interbase H bond can be given by an integration over the second derivative of a Morse potential weighted by a vibrational distribution function:

$$\phi_{li}^{\text{int}} = A_{li} \int_{r_{li}^{\min}}^{\infty} dr \exp\left[-\frac{(r - R_{li})^2}{2D_{li}} \right] \frac{d^2}{dr^2} V_{\text{Morse}}(r) \ . \quad (1)$$

The parameters of the Morse potential for both an AT pair and a GC pair can be found in Ref. [8]. We assume that these Morse parameters are independent of base sequence. In the above expression l and i are the index of base pairs in a unit cell and index of the interbase H bonds within a base pair, respectively. R_{li} is the mean bond length of a H bond which is a probability-weighted combination of mean intact bond length and open bond end-atom distance. The detailed formation of R_{li} will be given later in the paper. A_{li} is a normalization factor:

$$A_{li}^{-1} = \int_{r_{li}^{\min}}^{\infty} dr \exp[-(r - R_{li})^2 / 2D_{li}] \ . \quad (2)$$

$r_{li}^{\min}$ is the hard-core inner boundary and D_{li} is the mean vibrational amplitude given by

$$D_{li}=\frac{1}{\pi}\sum_{\lambda}\int_0^{\pi}d\theta\frac{|s_i^{\lambda}(\theta)|^2}{2\omega_{\lambda}(\theta)}\coth\left[\frac{\omega_{\lambda}(\theta)}{2k_BT}\right] . \tag{3}$$

Here θ is the phase angle, $\omega_{\lambda}(\theta)$ is the normal mode eigenfrequency of the system and $s_{li}^{\lambda}(\theta)$ is the projection of the difference of the H-bond end-atom eigenvectors onto the bond orientation.

In order to take into consideration the cooperative effect associated with the disrupted bonds and open base pairs one needs a bond breaking operator. In the case of cooperative MSPA this bond breaking operator is the individual bond disruption probability P_{li} and the base pair opening probability $P_l^{\rm op}$ defined by [7,8]

$$P_l^{\rm op}=\prod_i P_{li}\equiv\prod_i A_{li}\int_{L_{li}^{\max}}^{\infty}dr\exp[-(r-R_{li})^2/2D_{li}] , \tag{4}$$

where $L_{li}^{\max}$ is the maximum separation of the bond end atoms before disruption. The values of the $L_{li}^{\max}$ for both an AT pair and a GC pair are given in Ref. [8].

We have shown that [7,8] to exclude the disrupted bonds from contributing to the mean field calculation of MSPA one can use a probability-weighted linear combination of the intact bond force constant $\phi_{li}^{\rm int}$ and the disrupted bond force constant $\phi_{li}^{\rm dis}$ in the MSPA formalism in the following manner:

$$\phi_{li}=(1-P_{li})\phi_{li}^{\rm int}+P_{li}\phi_{li}^{\rm dis} , \tag{5}$$

where $\phi_{li}^{\rm dis}$ is the disrupted bond effective force constant and is set to zero.

The cooperative effect was also incorporated [7,8] into the MSPA by the use of a base pair opening probability-weighted linear combination in both the base stacking force constants and the interbase H-bond lengths that enter into the MSPA algorithm. However, to extend the theory to include heterogeneous base sequences the use of a single base pair open probability becomes insufficient. Unlike a mono-base-pair sequence where the probabilities $P_l^{\rm op}$ of different base pairs are the same, the probability $P_l^{\rm op}$ of a base pair in a heterogeneous sequence is expected to be significantly different from that of its neighbors. Therefore the base stacking force constants should be a function of two opening probabilities, i.e., the probabilities for the two adjoining base pairs whose stacking interaction is considered. On the other hand, to properly represent the constraint of a base pair by its two neighbors, the interbase H-bond lengths should be a function of three opening probabilities, one for the base pair considered and one for each of its two neighbors. Our analysis indicates that the best choice for the appropriate unstacking opening probability would be the geometric mean of the probabilities involved. This choice fits all the features required and it gives results in agreement with experimental observations as will be shown below. Therefore the effective base stacking force constant $\phi_{ll'}^{s}$ can be given by

$$\phi_{ll'}^{s}=(1-\sqrt{P_l^{\rm op}P_{l'}^{\rm op}})\phi_{ll'}^{\rm int}+\sqrt{P_l^{\rm op}P_{l'}^{\rm op}}\,\phi_{ll'}^{\rm op} , \tag{6}$$

where $\phi_{ll'}^{\rm int}$ and $\phi_{ll'}^{\rm op}$ are the base stacking force constants for intact and open base pairs, respectively. In practice $\phi_{ll'}^{\rm op}$ can be set to zero. The mean length of an interbase H bond can also be given by a weighted combination of intact bond length $R_{li}^{\rm int}$ and the unconstrained open bond end-atom distance $R_{li}^{\rm op}$. One can define a base pair unconstrained probability $P_l^{\rm uc}$ as

$$P_l^{\rm uc}=(P_{l-1}^{\rm op}P_l^{\rm op}P_{l+1}^{\rm op})^{1/3} \tag{7}$$

and the mean bond length can then be given by

$$R_{li}=(1-P_l^{\rm uc})R_{li}^{\rm int}+P_l^{\rm uc}R_{li}^{\rm op} . \tag{8}$$

The intact bond lengths and the open bond end-atom distances have been formulated previously. They are given by

$$\begin{aligned}R_{li}^{\rm int}&=r_{lu}^0+\frac{1}{a_{li}}\ln[\cosh(2a_{li}\sqrt{2D_{li}\ln 2})] ,\\ R_{li}^{\rm op}&=L_{li}^{\max}+2P_{li}\sqrt{2D_{li}} ,\end{aligned} \tag{9}$$

where r_{li}^0 and a_{li} are Morse parameters fitted to room-temperature data.

III. RESULTS AND DISCUSSION

We use the algorithm outlined in the preceding section to specifically calculate the opening probabilities of two heterogeneous sequence DNA copolymers at various temperatures. The two copolymers are the *B*-form poly[d(T-G)]·poly[d(C-A)] and the *B*-form poly[d(T-C)]·poly[d(G-A)]. For comparison we also include the results of another two *B*-DNA copolymers, poly[d(A-T)]·poly[d(A-T)] and poly[d(G-C)·poly[d(G-C)], as well as two *B*-DNA homopolymers, poly(dA)·poly(dT) and poly(dG)·poly(dC), from our earlier calculations [8]. The coordinates of these polymers are from the fiber studies [14]. Our calculation corresponds to a nominal salt concentration that is approximate to the number of experiments from which our parameters are obtained. This nominal salt concentration was estimated as $\sim 0.05M$ NaCl [7,15].

In our model, the nominal salt concentration is coupled with the refined parameters of the Morse potential that we use to describe the interbase H bonds. Our Morse parameters were fitted assuming that the H bonds were in equilibrium at the x-ray determined positions for the atoms involved at room temperature. These H bonds can be regarded as unstrained, i.e., the equilibrium bond lengths are determined by the particular Morse potentials and no further static force corrections are necessary at the nominal salt concentration. One can calculate the nominal salt concentration exactly as that concentration at which the salt shielded Coulomb strains across the H bonds from the phosphate groups are exactly canceled by other cross H-bond nonbonded interactions and the H-bond reaction forces at the x-ray determined configuration at room temperature. This also gives $0.05M$ NaCl [15] in agreement with the range of salt concentra-

tions reported for the samples used in the x-ray analysis and the Raman and ir analyses.

All calculations are for the *B* conformation, which is an approximation to the actual conformation the polymers are in. Our calculations strongly indicate the most important factors in base pair stability are, first, cross strand interactions, principally H bonds, and second, cross strand nonbonded forces. Small changes in conformation alter the normal mode distribution which enters into the stability calculation through the vibrational squared amplitude *D*. *D* integrates over the entire normal mode structure so that only changes in the mode structure alter this integral effect melting. We have carried out a calculation of the normal mode structure for poly(dA)·poly(dT) with propeller twist conformation alteration included [16]. Although specific differences in mode structure did occur, the integral effects of these shifts were very small. It seems unlikely that the bond disruption behavior will be strongly affected by minor changes in conformation.

A. Sequence dependence of the melting temperature

In Fig. 1 we show the calculated P_l^{op}*s* for the AT pairs at temperatures from room temperature (293 K) through the melting temperatures. Included in Fig. 1 are the probabilities of poly(dA)·poly(dT) with no addition of a minor groove spine of hydration (upper dashed line), poly(dA)·poly(dT) with addition of a minor groove spine of hydration as observed experimentally (lower dashed line), poly[d(A-T)]·poly[d(A-T)] (the upper solid line), the AT pair of poly[d(T-C)]·poly[d(G-A)] (the middle solid line) and the AT pair of poly[d(T-G)]·poly[d(C-A)] (lower solid line). The calculated P_l^{op}'s for the GC pairs are shown in Fig. 2. Included in Fig. 2 are the probabilities of poly(dG)·poly(dC) (dashed line), the GC pair of poly[d(T-C)·poly[d(G-A)] (upper solid line), the GC pair of poly[d(T-G)]·poly[d(C-A)] (middle solid line), and poly[d(G-C)]·poly[d(G-C)] (lower solid line).

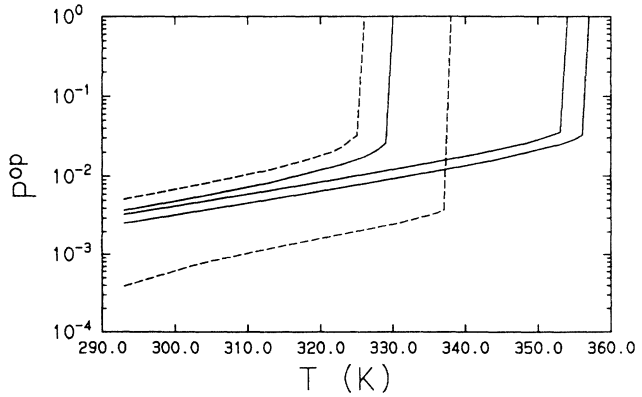

FIG. 1. Calculated base pair opening probability as a function of temperature for the AT base pair in the nonspine model of poly(dA)·poly(dT) (upper dashed line), poly(dA)·poly(dT) with the spine of hydration (lower dashed line), poly[d(A-T)]·poly[d(A-T)] (upper solid line), poly[d(T-C)·[d(G-A)] (middle solid line), and poly[d(T-G)]·poly[d(C-A)] (lower solid line).

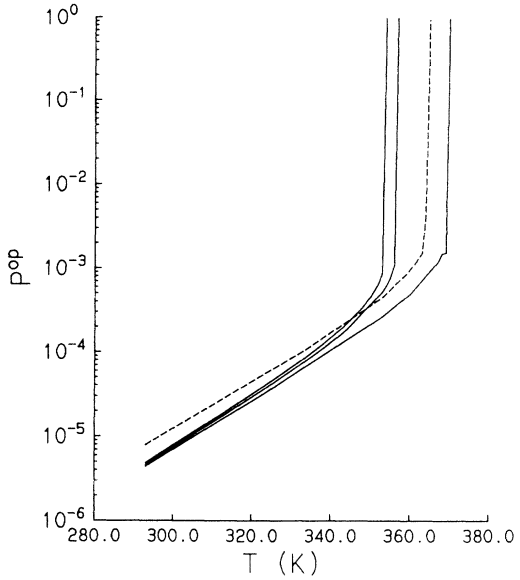

FIG. 2. Calculated base pair opening probability as a function of temperature for the GC base pair in poly(dG)·poly(dC) (dashed line), poly[d(T-C)]·poly[d(G-A)] (upper solid line), poly[d(T-G)]·poly[d(C-A)] (middle solid line), and poly[d(G-C)]·poly[d(G-C)] (lower solid line).

One finds from Figs. 1 and 2 that all curves show cooperative melting at a certain temperature such that the P^{op} rises sharply from $10^{-3}\sim10^{-2}$ to 1 consistent with observed behavior [17,18]. One can then define this critical transition temperature as the melting temperature of the DNA polymer. The calculated melting temperatures for the four copolymers and two homopolymers are given in Table I, together with experimentally measured melting temperatures at our nominal salt concentration [17,18]. As expected, our calculated T_m's of poly[d(T-C)]·poly[d(G-A)] and poly[d(T-G)]·poly[d(C-A)] are higher than those of mono-AT-pair polymers but lower than those of mono-GC-pair polymers, in agreement with observations. From Table I we also find that the calculated T_m's of these two copolymers as well as those of mono-base-pair sequence homopolymers and copolymers

TABLE I. Melting temperatures of *B*-DNA homopolymers and copolymers from our calculations (T_m^{theor}) and from experiments (T_m^{expt}). Experimental data are from Refs. [17,18].

Sample	T_m^{theory} (K)	T_m^{expt} (K)
poly(dA)·poly(dT) with no spine	326	
poly[d(A-T)]·poly[d(A-T)]	330	327
poly(dA)·poly(dT) with spine	338	335
poly[d(T-C)]·poly[d(G-A)]	354	350
poly[d(T-G)]·poly[d(C-A)]	357	356
poly(dG)·poly(dC)	365	367
poly[d(G-C)]·poly[d(G-C)]	370	378[a]

[a]Extrapolated from experimental data at salt concentrations lower than 0.05*M* NaCl by the following formula: $T_m = 396 + 14\log_{10}(C)$.

are also in fair agreement with the observed values.

Although both poly[d(T-C)]·poly[d(G-A)] and poly[d(T-G)]·poly[d(C-A)] are alternating AT-GC copolymers, our calculation as well as observations show that their melting temperatures are different. The melting temperature of the polymer with an alternating purine-pyrimidine sequence (poly[d(T-G)]·poly[d(C-A)]) is several degrees higher than its nonalternating counterpart (poly [d(T-C)]·poly[d(G-A)]). The same is true for the mono-base-pair polymers except in the case of poly(dA)·poly(dT) with a spine of hydration. For example, the melting temperature of poly[d(A-T)]·poly[d(A-T)] is higher than that of the non-hydration-spine model of poly(dA)·poly(dT). Similarly, the melting temperature of poly[d(G-C)]·poly[d(G-C)] is higher than that of poly(dG)·poly(dC). We have shown [8] that the reason the melting temperature of an alternating purine-pyrimidine sequence is higher than its nonalternating counterpart is because of the difference in the stacking interactions arising from the differences in the size of a purine base and that of a pyrimidine base. The size of a purine (adenine or guanine) base is substantially larger than a pyrimidine (thymine or cytosine) base. Because of substantially larger sizes of the purine bases, the overlap between the purine bases in an alternating purine-pyrimidine sequence for every second base pair should be significantly greater than the overlapping between a purine base and a pyrimidine base in a nonalternating purine-pyramidine sequence. Therefore one expects that the cross strand base stacking in an alternating purine-pyrimidine sequence is stronger than its nonalternating counterpart. A polymer with an alternating purine-pyrimidine sequence should be more stable than its nonalternating counterpart. Indeed this is confirmed both by our calculation and by experimental observations.

Based on the same argument one would also expect that the melting temperature differences between polymers containing GC pairs should be larger than those containing AT pairs. The reason can be given as follows. From Fig. 3 one can see that although the size of an adenine base is similar to the size of a guanine base, the guanine base has an additional O(6) atom on the floor of the major groove. This atom is on the inner side of the base and forms an interbase H bond to the cytosine amino group on the opposite strand. Because of its favorable position, this additional atom should contribute an additional cross strand base stacking force. This additional cross strand base stacking force would therefore contribute to an additional stability to the alternating GC sequences. Indeed our calculation shows that the melting temperature difference between poly[d(G-C)]·poly[d(G-C)] and poly(dG)·poly(dC) is greater than the difference between poly[d(A-T)]·poly[d(A-T)] and the non-hydration-spine model of poly(dA)·poly(dT).

From Table I one finds that the calculated melting temperature difference between poly[d(T-G)]·poly[d(C-A)] and poly[d(T-C)]·poly[d(G-A)] is smaller than the difference between poly[d(A-T)]·poly[d(A-T)] and the non-hydration-spine model of poly(dA)·poly(dT). This at first seems to be in contradiction with the above argument. However, a closer examination of the structures of AT and GC pairs from Fig. 3 shows that the smaller difference seen in the alternating AT-GC copolymers is a result of a stronger guanine-thymine stacking than adenine-thymine stacking coupled by a weaker increase in the stacking between guanine and adenine with respect to that between adenine and adenine. From Fig. 3 one can see that unlike the guanine base the adenine base does not have an amino group in the major groove. As a result the increase in the stacking between a guanine base and an adenine base with respect to that between two adenine bases is not as strong as that between two guanine bases. On the other hand, the thymine base has an O(2) atom exposed on the floor of the major groove and on the edge close to the opposite strand. As a result the cross strand stacking interaction between a guanine base and a thymine base is expected to be stronger than the interaction involving adenine base and thymine base. Therefore a less significant increase in the stacking stability in poly[d(T-G)]·poly[d(C-A)] with respect to that in poly[d(A-T)]·poly[d(A-T)] coupled with an increase in the stacking stability in poly[d(T-C)]·poly[d(G-A)] with respect to that in the non-hydration-spine model of poly(dA)·poly(dT) may contribute to the smaller melting temperature difference between poly[d(T-G)]·poly[d(C-A)] and poly[d(T-C)]·poly[d(G-A)].

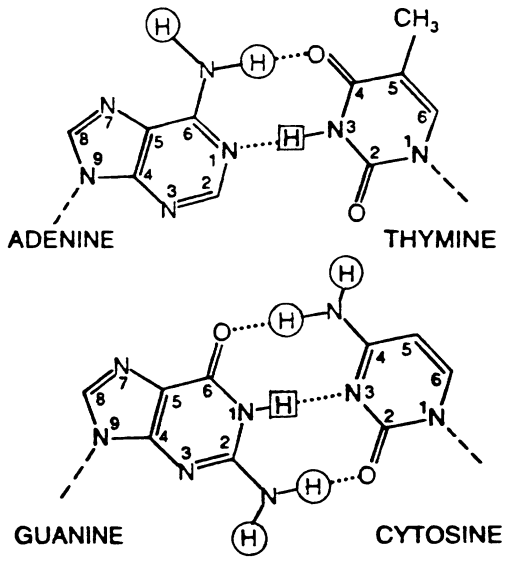

FIG. 3. Structure of both AT and GC base pairs. The sugar-phosphate backbone is not shown. The exchangeable amino protons are circled and the imino protons are enclosed in a square.

B. Sequence effect in the premelting regime

In Tables II and III we give AT and GC pair opening probabilities for the polymers studied at several premelting temperatures together with some experimental estimates [19–24]. Our calculation indicates that, except for

TABLE II. Premelting AT base pair opening probability of (A) nonspine model of poly(dA)·poly(dT), (B) poly(dA)·poly(dT) with a minor groove spine of hydration, (C) poly[d(A-T)]·poly[d(A-T)], (F) poly[d(T-C)]·poly[d(G-A)], and (G) poly[d(T-G)]·poly[d(C-A)], (Expt. 1) experimental estimate for AT (AU) pairs without the spine, and (Expt. 2) experimental estimate for AT pairs with the spine. Most experiments were carried out at about 298 K and at a slightly different salt concentration than 0.05*M* NaCl (details in Refs. [20,21] and references in Refs. [19,22]).

	P^{op}						
T (K)	(A)	(B)	(C)	(F)	(G)	Expt. 1	Expt. 2
293	0.005 17	0.000 39	0.003 76	0.003 34	0.002 59	$\sim 10^{-3}$	$\sim 10^{-5}$
313	0.01197	0.001 18	0.008 23	0.006 61	0.005 11		
323	0.023 27	0.001 85	0.013 85	0.009 36	0.007 21		
333		0.002 91		0.013 46	0.010 31		
353				0.035 87	0.025 03		

the AT pairs in poly(d(A)·poly(dT) with a minor groove spine of hydration, base sequence has a definite but rather limited effect on the open pair probability at temperatures far away from the helix-coil transition region. We find that both the differences in the opening probabilities for the AT pairs and the differences for the GC pairs of those polymers except the polymer with the hydration spine are quite small at these premelting temperatures even though the melting temperatures are significantly different. The base pair in poly(dA)·poly(dT) with a hydration spine is an exception because the spine enhances the stability of the base pair and as a result the opening probability is reduced significantly [19]. Our calculation shows that the P^{op} of an alternating purine-pyrimidine sequence is smaller than that of its nonalternating counterpart because of strong cross strand stacking interactions in the former sequence as discussed earlier. However, the difference is quite small; it is less than 30% for the AT pairs and less than 50% for the GC pairs. The P^{op} of an AT pair in an alternating AT-GC sequence is smaller than that in a mono-base-pair sequence because of the enhanced stability contributed by the GC pairs. On the other hand, the P^{op} of a GC pair in an alternating AT-GC sequence is larger than that in the GC copolymer because of reduced stability contributed by the AT pairs. Again the difference is small; it is less than 50% for the AT pairs and less than 40% for the GC pairs at temperatures below the transition region.

We find that the P^{op} of a GC pair in poly(dG)·poly(dC) is slightly larger than that of a GC pair in poly[d(T-C)]·poly[d(G-A)] and poly[d(T-G)]·poly[d(C-A)] at temperatures below 345 and 349 K, respectively. Our analysis indicates that this inverted stability arises because of an additional softening to the base pair in the homopolymer coupled by a weaker cooperativity in the copolymers in this temperature range. The additional softening arises from the differences in the distribution of the interbase H-bond soft modes between poly(dG)·poly(dC) and the AT-GC copolymers. A number of soft modes in poly(dG)·poly(dC) are in a lower-frequency range than the soft modes in the two copolymers. As a result the number of phonons ($\approx kT/\hbar\omega$) of these modes is higher. This extra number of soft mode phonons contributes to an additional softening to the GC pair in the homopolymer. This additional softening seems to be stronger than the elements that contribute to the cross strand stacking between GC pairs in this temperature range. On the other hand, the destabilizing effect of the AT pairs on a GC pair in the AT-GC copolymers is small and it is not enough to soften the GC pairs significantly because of low cooperativity in this temperature range. This destabilizing effect as well as the cross strand base stacking interactions become more and more important at higher temperatures. At higher temperatures these factors outweigh the effects associated with the differences in the soft mode distribution. As shown in Fig. 2, the P^{op} of a GC pair in both poly[d(T-C)]·poly[d(G-A)] and poly[d(T-G)]·poly[d(C-A)] becomes

TABLE III. Premelting GC base pair opening probability of (D) poly(dG)·poly(dC), (E) poly[d(G-C)]·poly[d(G-C)], (F) poly[d(T-C)]·poly[d(G-A)], (G) poly[d(T-G)]·poly[d(C-A)], and (Expt.) experimental estimate. Most experiments were carried out at about 298 K and at a slightly different salt concentration than 0.05*M* NaCl (details in Ref. [24] and references in Ref. [23]).

	P^{op}				
T (K)	(D)	(E)	(F)	(G)	Expt.
293	7.91×10^{-6}	4.40×10^{-6}	4.87×10^{-6}	4.60×10^{-6}	$\sim 10^{-6}$
313	2.81×10^{-5}	1.63×10^{-5}	1.91×10^{-5}	1.78×10^{-5}	
323	5.49×10^{-5}	3.13×10^{-5}	3.87×10^{-5}	3.55×10^{-5}	
333	1.02×10^{-4}	6.14×10^{-5}	8.23×10^{-5}	7.41×10^{-5}	
353	4.47×10^{-4}	2.64×10^{-4}	8.64×10^{-4}	5.19×10^{-4}	

TABLE IV. Premelting adenine amino interbase H-bond [N(6)—H—O(4) bond] disruption probability P_{Aam} of an AT pair in (A) nonspine model of poly(dA)·poly(dT), (B) poly(dA)·poly(dT) with a minor groove spine of hydration, (C) poly[d(A-T)]·poly[d(A-T)], (F) poly[d(T-C)]·poly[d(G-A)], (G) poly[d(T-G)]·poly[d(C-A)], and (Expt.) experimental estimate (Ref. [27]).

	P_{Aam}					
T (K)	(A)	(B)	(C)	(F)	(G)	Expt.
293	0.0442	0.0404	0.0449	0.0404	0.0412	0.05
313	0.0772	0.0632	0.0748	0.0624	0.0634	
323	0.1218	0.0796	0.1058	0.0781	0.0791	
333		0.1012		0.0991	0.0998	
353				0.1917	0.1889	

larger than that of the GC pair in poly(dG)·poly(dC) at temperatures higher than 345 and 349 K, respectively. From Figs. 1 and 2 one can also see that the differences in the opening probabilities become large in the region close to the critical transition temperatures.

Premelting base pair opening probability can be measured experimentally by imino proton exchange [20,21]. As shown in Fig. 3, the imino proton is buried inside the double helix. Therefore the exchange of this proton with solvent can only occur in an all H-bond disrupted base pair open state [22,23]. Imino proton exchange measurements at room temperatures [20,21,25] showed that, for base pairs in the central region of an oligomer where end effects are small, the exchange rate as well as the opening probabilities are determined mainly by the nature of the base pair. Although there is a certain sequence dependence for the base pair lifetime, this dependence is only significant for those base sequences such as the long A tract which have a rather narrow minor groove. It has been shown that a spine of hydration can be formed in a narrow minor groove of a random DNA sequence [26]. Therefore the anomalously long base pair lifetime seen in the long A tracts is likely to correlate with the enhanced stability caused by the spine. These observed features are all seen in our calculated probabilities.

We have shown [19,22,23] that our calculated base pair opening probabilities for both AT and GC pairs are in fair agreement with various experimental measurements. Moreover, our calculated individual bond disruption probabilities are also in fair agreement with amino proton exchange measurements [22,23]. Amino proton exchange measures the exchange of the two protons in the amino group with the solvent [27,28]. As seen from Fig. 3, since one of the hydrogen atoms is involved in the interbase H bonding, the exchange requires some sort of open state to facilitate the process. We have shown [22,23] that the breakdown of an amino interbase H bond (N—H—O bond) can be associated with the "open state" needed to facilitate the corresponding amino proton exchange. The calculated amino interbase H-bond disruption probabilities for the AT pairs and GC pairs are given in Tables IV and V, respectively. In Table IV we give an available experimental estimate of the bond disruption probability of the adenine amino H bond.

C. Analysis of the formulation of the near-neighbor cooperative effect

We have formulated the near-neighbor cooperative effect in a heterogeneous DNA sequence by the use of the geometric mean of the base pair opening probabilities as the cooperative element. To examine how sensitively our calculation is dependent on this formulation, we carry out a calculation where a different cooperativity is used. We choose to use $P_l^{\text{uc}}=P_l^{\text{op}}$ in this calculation. This approaches our realistic cooperativity if it is used in mono-base-pair polymers since in that case the probabilities of different base pairs are the same. For a heterogeneous sequence the P_l^{op}'s for various base pairs differ even at temperatures well below the melting temperature. By using this unrealistic formulation only the information of a sin-

TABLE V. Premelting amino interbase H-bond disruption probabilities of a GC base pair in (D) poly(dG)·poly(dC), (E) poly[d(G-C)]·poly[d(G-C)], (F) poly[d(T-C)]·poly[d(G-A)], (G) poly[d(T-G)]·poly[d(C-A)], and (Expt.) experimental estimate (carried out at 273 K, for details see Ref. [28] and references in Ref. [23]). P_{Cam} is the disruption probability of the cytosine amino H bond [N(4)—H—(6) bond] and P_{Gam} is the disruption probability of the guanine amino H bond [O(2)—H—N(2) bond].

	P_{Cam}					P_{Gam}				
T (K)	(D)	(E)	(F)	(G)	Expt.	(D)	(E)	(F)	(G)	Expt.
293	0.0057	0.0051	0.0051	0.0051	0.005	0.032	0.023	0.026	0.025	0.05
313	0.0105	0.0095	0.0098	0.0096		0.049	0.036	0.041	0.039	
323	0.0142	0.0129	0.0136	0.0133		0.062	0.045	0.052	0.050	
333	0.0190	0.1078	0.0193	0.0188		0.075	0.056	0.067	0.064	
353	0.0372	0.0354	0.0568	0.0465		0.126	0.090	0.149	0.122	

gle base pair is used as the cooperative element. The element associated with a base pair being constrained by its neighbors is not properly represented. Apart from P_l^{uc} no other changes are made in the calculation. Our calculation is carried out on poly[d(T-G)]·poly[d(C-A)] and the calculated P^{op}'s for both the AT pair (upper dashed line) and the GC pair (lower dashed line) as a function of temperature are shown in Fig. 4. For comparison we also include in Fig. 4 the P^{op}'s of both the AT pair (upper solid line) and GC pair (lower solid line) of poly[d(T-G)]·poly[d(C-A)] calculated by the use of our realistic P_l^{uc}.

We find from Fig. 4 that at temperatures well below the AT polymer melting temperature (330 K) the differences in the calculated probabilities for both the AT pair and the GC pair using the two algorithms are actually very small. This is understandable because in this temperature regime the cooperative effect is very weak because of the low opening probabilities of both AT and GC pairs. As a result, the calculation is insensitive to the details of the formulation of the cooperative effect. The difference in the P^{op}'s for the AT pairs becomes significantly larger at temperatures immediately above 330 K, while the difference in the GC pairs is still very small in this temperature range. The calculation using the unrealistic cooperativity would predict the melting of all the AT pairs at 339 K without the large-scale melting of the GC pairs. The GC pairs only melt at 364 K, which is 8 K above the observed melting temperature. The melting of AT pairs without the melting of GC pairs in an alternating AT-GC sequence shown here is a direct result of the misrepresentation of the element associated with the constraint of a base pair by its neighbors. This behavior is, however, not the observed behavior. Experiments [18] as well as our calculation using realistic cooperativity show that because of cooperative effects large-scale melting of both the AT pairs and GC pairs in an alternating AT-GC sequence occurs at a single temperature in spite of the significant difference in the stability between an AT pair and a GC pair.

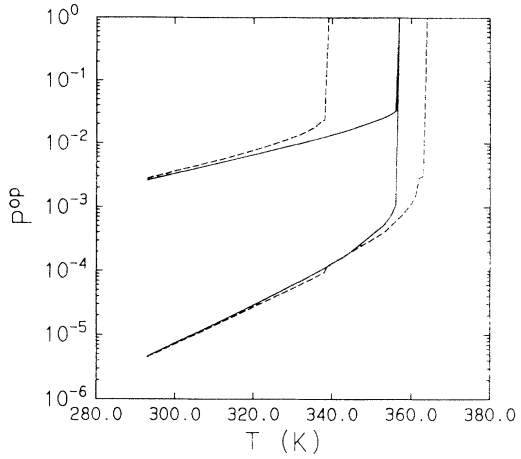

FIG. 4. Calculated base pair opening probabilities of poly[d(T-G)]·poly[d(C-A)] using different description of cooperativity. The solid lines correspond to the calculation using $P_l^{uc}=(P_{l-1}^{op}P_l^{op}P_{l+1}^{op})^{1/3}$ and the dashed lines correspond to the calculation using $P_l^{uc}=P_l^{op}$. The upper solid (dashed) line is for the AT pair and the lower solid (dashed) line is for the GC pair.

ACKNOWLEDGMENT

This work was supported in part by ONR Contract No. N00014-92-K-1232.

[1] D. Poland, Biopolymers **13**, 1859 (1974).
[2] M. Fixman and J. J. Freire, Biopolymers **16**, 2693 (1977).
[3] A. V. Vologodskii, B. R. Amirikyan, Y. L. Lyubcheko, and M. D. Frank-Kamenetskii, J. Biomol. Struct. Dyn. **2**, 131 (1984).
[4] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
[5] Y. Gao, K. V. Devi-Prasad, and E. W. Prohofsky, J. Chem. Phys. **80**, 6291 (1984).
[6] E. W. Prohofsky, in *Biomolecular Stereodynamics IV*, Proceedings of the 4th Conversation in the Discipline of Biomolecular Stereodynamics, edited by R. H. Samar and M. H. Samar (Adenine, New York, 1985), pp. 21–46.
[7] Y. Z. Chen and E. W. Prohofsky, Biopolymers **33**, 351 (1993).
[8] Y. Z. Chen and E. W. Prohofsky, Biopolymers **33**, 797 (1993).
[9] N. R. Werthamer, Phys. Rev. B **1**, 572 (1970).
[10] M. Tsuboi, S. Takahashi, and I. Harada, in *Physico-Chemical Properties of Nucleic Acids*, edited by J. Duchesne (Academic, New York, 1973), Vol. 2, pp. 91–145.
[11] K. C. Lu, E. W. Prohofsky, and L. L. Van Zandt, Biopolymers **16**, 2491 (1977).
[12] W. N. Mei, M. Kohli, E. W. Prohofsky, and L. L. Van Zandt, Biopolymers **20**, 833 (1981).
[13] V. V. Prabhu, L. Young, E. W. Prohofsky, and G. S. Edwards, Phys. Rev. B **39**, 5436 (1989).
[14] R. Chandrasekaran and A. Arnott, in *The Structures of DNA and RNA Helices in Oriented Fibres*, edited by W. Saenger, Landolt-Börnstein, Numerical Data and Functional Relationships in Science and Technology Vol. VII/1b (Springer-Verlag, Berlin, 1989), pp. 31–170.
[15] Y. Z. Chen and E. W. Prohofsky, Phys. Rev. E **48**, 3099 (1993).
[16] L. Young, V. V. Prabhu, and E. W. Prohofsky, Phys. Rev. A **41**, 7020 (1990).
[17] R. B. Inman and R. L. Baldwin, J. Mol. Biol. **8**, 452 (1964).
[18] R. D. Wells, J. E. Larson, R. C. Grant, B. E. Shortle, and C. R. Cantor, J. Mol. Biol. **54**, 465 (1970).
[19] Y. Z. Chen and E. W. Prohofsky, Nucleic Acids Res. **20**, 415 (1992).
[20] M. Gueron, M. Kochoyan, and J.-L. Leroy, Nature (London) **328**, 89 (1987).
[21] M. Gueron, E. Charretier, J. Hagerhorst, M. Kochoyan, J. L. Leroy, and A. Moraillon, in *Structure & Methods, Vol. 3: DNA & RNA*, edited by R. H. Sarma and M. H. Sarma

(Adenine, New York, 1990), pp. 113–137.
[22] Y. Z. Chen, Y. Feng, and E. W. Prohofsky, Biopolymers **31**, 139 (1991).
[23] Y. Z. Chen, W. Zhuang, and E. W. Prohofsky, Biopolymers **31**, 1273 (1991).
[24] M. Kochoyan, J. L. Leroy, and M. Gueron, Biochemistry **29**, 4799 (1990).
[25] M. Leijon and A. Graslund, Nucleic Acids Res. **20**, 5339 (1992).
[26] V. P. Chuprina, U. Heinemann, A. A. Nurislamov, P. Zielenkiewicz, R. E. Dickerson, and W. Saenger, Proc. Natl. Acad. Sci. U.S.A. **88**, 593 (1991).
[27] R. S. Preisler, C. Mandal, S. W. Englander, and N. R. Kallenbach, Biopolymers **23**, 2099 (1984).
[28] H. Teitelbaum and S. W. Englander, J. Mol. Biol. **92**, 79 (1975).