Modified self-consistent phonon calculation of the dependence of DNA melting temperature on guanine-cytosine content

Y. Feng and S. C. Tan
Department of Physics, National University of Singapore, Singapore 0511, Singapore
(Received 18 August 1992)

We have modified the self-consistent phonon approach to treat DNA double helices modeled with a large repeating unit (N base pairs) so that melting behaviors of DNA polymers of various base-pair sequences can be studied. The melting temperatures of DNA polymers of different G-C–to–A-T ratios were calculated (G, C, A, and T denote guanine, cytosine, adenine, and thymine, respectively). It is shown that a DNA polymer with a relatively higher content of A-T melts at a lower temperature than DNA with a higher content of G-C. The melting temperature increases linearly when the G-C content in a DNA polymer increases. This is in agreement with both experimental observations and theoretical analyses.

PACS number(s): 87.15.By, 63.20.Dj, 63.70.+h, 64.60.Ht

I. INTRODUCTION

It is known that the stability of natural DNA depends on many factors such as the size, base-pair sequence of the molecule, and ionic strength of the medium [1]. The stability of a DNA double helix is characterized by the melting temperature at which the helix-coil transition takes place. For a single long DNA molecule, the melting temperature depends mainly on the guanine-cytosine (G-C) and adenine-thymine (A-T) base-pair ratio. As the G-C base pair is more stable than the A-T base pair, a DNA polymer with a relatively higher content of G-C melts at a higher temperature than DNA with a higher content of A-T. To a good approximation, the melting temperature increases linearly when the G-C content in DNA polymer increases [2].

Because of important functions played by DNA denaturation in replication, transcription, and recombination, many theoretical and experimental techniques have been developed to study melting or strand separation of DNA. One of the successful theoretical analyses is the modified self-consistent phonon approach [3] (MSPA), which provides a fundamental understanding of the melting process based on microscopic motions of atoms. The method has been used to predict melting temperatures for a number of DNA homopolymers and copolymers [4–6]. The predicted melting temperatures are close to observed values and rank the melting behavior in the proper order for these materials. The theory was also generalized to calculations which attempted to simulate melting of a DNA helix from an initiation or nucleation site where a defect was introduced, and to simulate the growth of melted defects [7–10]. The melting temperatures found are essentially characteristic of these materials. Further studies also showed that the MSPA gives fairly good estimates to premelting thermal fluctuational base-pair opening probabilities [11, 12].

Due to limitations of resources such as computing power, however, only a limited number of different repeating polymers which can be made with repeating units no larger than two base pairs were studied in previous calculations [4–6]. These include only a few possible DNA double helices [1, 6] and three possible values of G-C contents: 0% [poly(dA)-poly(dT)], 50% [poly(dGA)-poly(dTC) or poly(dGT)-poly(dAC)], and 100% [poly(dG)-poly(dC)]. In this study, we extend the MSPA calculation to DNA polymers containing a number of different G-C–to–A-T ratios. This is done by choosing a larger repeating unit, the unit cell, and varying nucleotide compositions in the unit cell. Calculations are carried out to analyze the dependence of melting temperature on G-C content. While the relationship between DNA melting temperature and G-C content has been well established, the current study would provide a further verification of the linear relationship. More importantly, it would serve as a further test to the validity of the MSPA theory.

II. THE MSPA

The MSPA formalism has been described in a number of publications [4–6, 13, 14]. The approach is based on the lattice-dynamic theory which was devised for the study of crystalline solids. A repeating DNA double helix can be considered as a system extended in one dimension. The helical symmetry translates whatever wave function existing at one region to another region a distance a away. This enables the equation of motion to be reduced to one associated with the degree of freedom of a unit cell. The primitive repeating cell is one consisting only one base pair. This gives a DNA double helix of uniform base pairs, or homopolymer. For such a double helix, there are 41 atoms in a unit cell in an approximation in which hydrogen atoms are attached to the nearest heavy atoms. A normal-mode analyses would produce wavelike vibrational excitations and these wavelike solutions form 123 bands.

A variety of DNA double helices can be modeled by choosing a larger repeating unit and varying the nu-

cleotide composition of the repeating unit. However, the required computing resources would increase dramatically as a result of increase in size of the unit cell. For example, choosing a unit cell consisting of two base pairs [6], the number of coupled equations that must be solved simultaneously becomes 246 compared to 123 for a homopolymer. And solving a system of 246 coupled equations would require much more than double the efforts in solving a system of half of its size. Up to now, the largest unit cell used has been two base pairs in previous MSPA calculations.

With numerous fast speed computers available today, however, it becomes possible to perform a normal-mode analyses for DNA double helices made with repeating units of more than two base pairs. With a larger unit cell, a large number of DNA polymers can be sampled with a variety of repeating units. One may arrange the base-pair compositions of the repeating units so that more realistic DNA polymers can be simulated and make it possible to compare theoretical predictions directly with experimental observations. In this study, we extend the MSPA formalism to cater a DNA double helix made with a repeating unit of N base pairs, where N is any integral number. Using the same approximation of combining hydrogen atoms with the heavy atoms with which they attach, there would be $41N$ atoms in a unit cell. One now has to solve a system of $123N$ coupled equations, or

$$(\underline{F} - \omega^2 \underline{I})\underline{\chi} = \underline{0} \tag{1}$$

in matrix form. In Eq. (1), $\underline{F}$ is the dynamic matrix ($123N \times 123N$), ω is the eigenvalue, $\underline{\chi}$ is the eigenvector, and $\underline{I}$ is a unitary matrix. The solutions of Eq. (1) form $123N$ bands for a long helix.

In the MSPA approach, the hydrogen bonds which link the two complementary strands are modeled by the Morse potential [15],

$$V(R) = V_0[1 - e^{-a(R-R_0)}]^2 - V_0. \tag{2}$$

When temperature increases, the hydrogen bonds are found to soften and eventually break at the melting temperature of the double helix. The mean-squared vibrational amplitude and the anharmonic force constant for the ith hydrogen bond are obtained from [5]

$$D_i = \frac{1}{\pi} \sum_{\lambda} \int_0^{\pi} d\theta \frac{|s_i^{\lambda}(\theta)|^2}{2\omega_{\lambda}(\theta)} \coth\left[\frac{\hbar\omega_{\lambda}(\theta)}{2kT}\right] \tag{3}$$

and

$$\phi_i = \frac{\int du e^{-u^2/2D_i} \frac{d^2V_i[\bar{R}_i(T)+u]}{du^2}}{\int du e^{-u^2/2D_i}}, \tag{4}$$

where $s_i^{\lambda}(\theta)$ is the projection of the λth normal-mode oscillation to the ith hydrogen bond stretch and $\bar{R}_i(T)$ is the average equilibrium length of the ith hydrogen bond at temperature T, which is determined by

$$V[\bar{R}_i(T) + \mu(T)] = V[\bar{R}_i(T) - \mu(T)], \tag{5}$$

if $\mu(T)$ is the amplitude of oscillation of the thermal phonons.

We followed the same procedures as described in previous publications [4–6, 14] in calculating the melting temperature of a DNA double helix. Because of softening in hydrogen bonds, Eqs. (1), (3), and (4) must be solved self-consistently using an iterative method at a given temperature. Then the temperature is incremented and the process is repeated until a melting criterion is satisfied. The corresponding temperature is defined as the melting temperature of the DNA helix. It has been shown [14, 16] that fairly accurate vibrational amplitudes and anharmonic force constants can be obtained for the hydrogen bonds involved using the Green's function and perturbation technique [5] for temperatures below and up to the melting temperature while saving computing time compared to direct matrix diagonalization at each iteration. We therefore used the same technique in the present study. Using this technique, the difference between the anharmonic force constant obtained from Eq. (4) and the initial quasiharmonic force constant used in Eq. (1) to obtain the normal-mode frequencies and eigenfunctions is treated as a perturbation.

To study the relationship between melting temperature and the G-C content of DNA polymers, we have carried out calculations for DNA double helices formed with repeating units up to five base pairs. The base pairs of

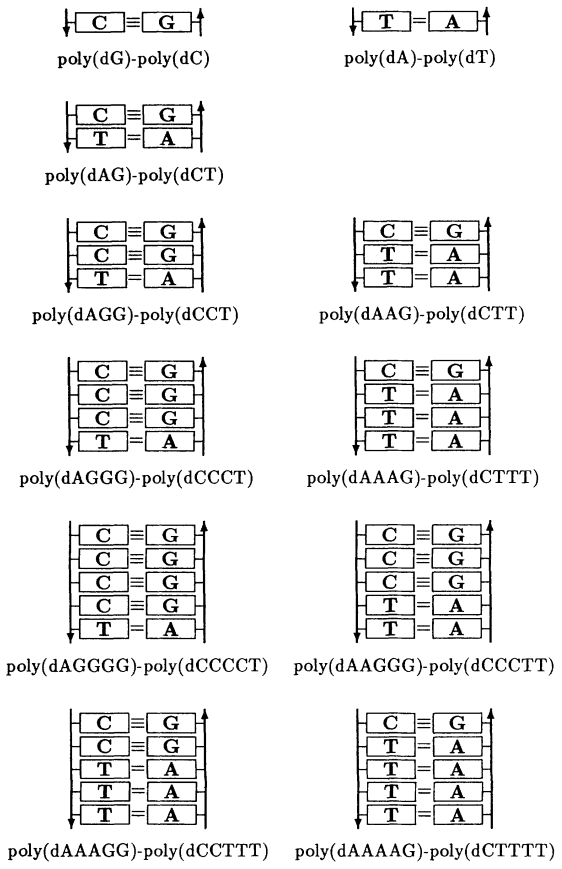

FIG. 1. Repeating units for DNA double helices being studied.

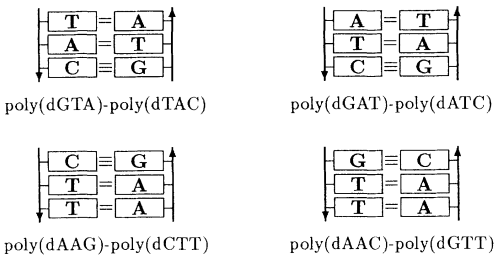

FIG. 2. Examples of repeating units of DNA double helices with the same G-C–to–A-T ratio but different base-pair arrangements.

these DNA double helices are simply arranged as shown in Fig. 1.

One may also arrange the base-pair sequence in various patterns to form a number of different DNA polymers for a given G-C–to–A-T ratio to study the dependence of melting temperature on base-pair sequence. Some of the DNA double helices which can be formed by choosing a repeating unit of three base pairs are shown in Fig. 2. All these DNA helices have the same G-C content, 33.3%. However, their melting temperatures could be different. The B conformation is assumed for all DNA helices being studied.

III. CRITERION FOR DNA MELTING

As temperature increases, the self-consistent bond length and vibrational amplitude increase while the anharmonic force constant decreases for each hydrogen bond due to softening in the hydrogen bond. In previous calculations [9,16], we used the turning point in the average potential energy as a criterion for melting. The drop in the average potential is an indication of non-solid-like behavior and is believed to be associated with initiation of melting of the system. Further study [17] showed that the drop in average potential is always accompanied by a rapid decrease in the anharmonic force constant or rapid increases in vibrational amplitude and thermal expansion in the hydrogen bonds and they all are indications of initiation of helix melting. Therefore, we may, in principle, use any of these phenomena as a criterion for helix melting. In this study, we used a melting criterion which is based on thermal expansions in hydrogen bonds.

The complementary strands are linked by relatively weak hydrogen bonds. As temperature increases, interstrand breathing vibrations cause the hydrogen bonds to be stretched. This thermal expansion is taken into account in the MSPA treatment via the nonlinear Morse potential. Since a self-consistent solution can be obtained at any given temperature with a fairly accurate computing procedure [14], we must therefore assume that the helix melts when the thermal expansion reaches a certain critical value. After analyzing a number of different DNA polymers and comparing with experimentally observed melting temperatures as well as previous calculations, we used the following as the melting criterion in the present study,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\bar{R}_i(T)-\bar{R}_i(T_0)}{\bar{R}_i(T_0)} > 1\%, \tag{6}$$

where T_0 is room temperature and n is the number of hydrogen bonds in a unit cell. The quantity on the left-hand side of (6) represents an average of relative increases in average hydrogen bond lengths from the equilibrium values at room temperature. Using this criterion, the calculated melting temperatures for poly(dA)-poly(dT) and poly(dG)-poly(dC) in the B conformation are 328 K and 365 K, respectively, being consistent with previous calculations. The melting temperatures are slightly higher than the experimentally observed values because the MSPA is a mean-field theory which assumes that anything that happens in one unit cell would happen in all other unit cells. While in reality, the melting is almost always initiated at a specific region and the melted region then propagates along the helix, which requires less amount of energy.

IV. RESULTS

Using the MSPA approach, we have calculated the melting temperatures of the 11 DNA double helices shown in Fig. 1, which represent 11 different G-C contents: 0%, 20%, 25%, 33.3%, 40%, 50%, 60%, 66.7%, 75%, 80%, and 100%. The calculated melting temperatures are listed in Table I and are also presented in Fig. 3. As expected, our results show that a DNA polymer with a higher content of G-C melts at a higher temperature than a DNA polymer with a lower G-C content. The melting temperature increases linearly as the G-C content increases, in agreement with experimental observations and other theoretical analyses.

A DNA polymer with a higher G-C content melts at a higher temperature than a DNA helix with a lower G-C content because the G-C base pair is more stable than the A-T base pair. The guanine and cytosine bases are bridged by three hydrogen bonds while the adenine and thymine bases are linked by two hydrogen bonds. In a DNA helix with mixed G-C and A-T base pairs, the A-T base pairs are stabilized by the G-C base pairs and the G-

TABLE I. G-C contents and calculated melting temperatures for the 11 DNA double helices shown in Fig. 1.

DNA	G-C (%)	T_m (K)
poly(dA)-poly(dT)	0	328
poly(dAAAAG)-poly(dCTTTT)	20	332
poly(dAAAG)-poly(dCTTT)	25	335
poly(dAAG)-poly(dCTT)	33.3	339
poly(dAAAGG)-poly(dCCTTT)	40	340
poly(dAG)-poly(dCT)	50	345
poly(dAAGGG)-poly(dCCCTT)	60	346
poly(dAGG)-poly(dCCT)	66.7	350
poly(dAGGG)-poly(dCCCT)	75	353
poly(dAGGGG)-poly(dCCCCT)	80	355
poly(dG)-poly(dC)	100	365

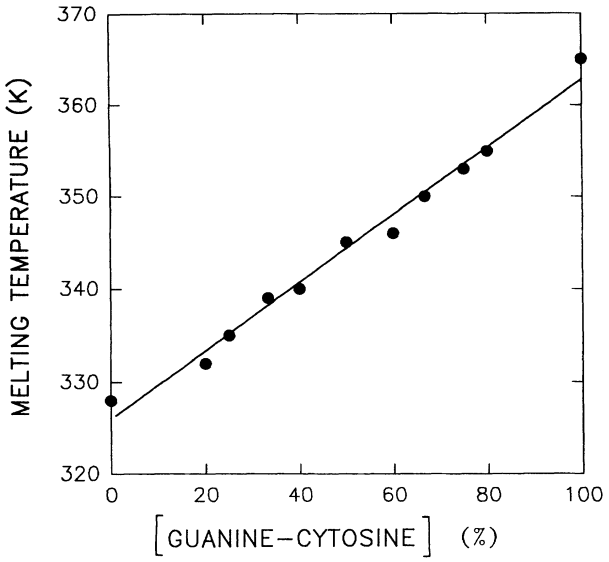

FIG. 3. Dependence of the calculated melting temperature on G-C content.

C base pairs are destabilized by the A-T base pairs. The stabilization is more significant if a helix contains more G-C base pairs while the destabilization is more significant in a helix with more A-T base pairs. This is shown in Figs. 4 and 5 by the larger average hydrogen bond lengths in G-C base pairs in a helix with a higher A-T content and the smaller average hydrogen bond lengths in A-T base pairs in a helix with a relatively higher G-C content in our calculation. In Fig. 4, the thermal average of hydrogen bond length is shown as a function of temperature for the major and minor groove hydrogen bonds in the A-T base pairs of poly(dA)-poly(dT), poly(dAGGG)-poly(dCCCT), and the middle A-T base pair of poly(dAAAG)-poly(dCTTT). It is obvious that the average hydrogen bond length is reduced in a helix with higher G-C content. In Fig. 5, the same quantity is shown for the major and minor groove bonds in the G-C base pairs of poly(dG)-poly(dC), poly(dAAAG)-poly(dCTTT), and in the middle G-C base pair of poly(dAGGG)-poly(dCCCA). The average bond length is larger in a helix with higher content of A-T base pairs due to destabilization by the A-T base pairs. Similar results are observed for the middle hydrogen bonds in G-C base pairs and hydrogen bonds in all other base pairs. They are not shown in the figures for clarity.

For DNA helices of the same G-C content, the stability depends on the actual base-pair sequence. For example, poly(dAAAGG)-poly(dCCTTT) might melt at a lower temperature than poly(dAGAGA)-poly(dTCTCT) due to the fact that the A-T base pair in the middle of the dAAAGG block does not have G-C neighbors and the destabilization by the G-C base pairs is less significant than those in poly(dAGAGA)-poly(dTCTCT). Even for the same A-T and G-C sequence, a simple flip over of any base pair could result in a change in stability of

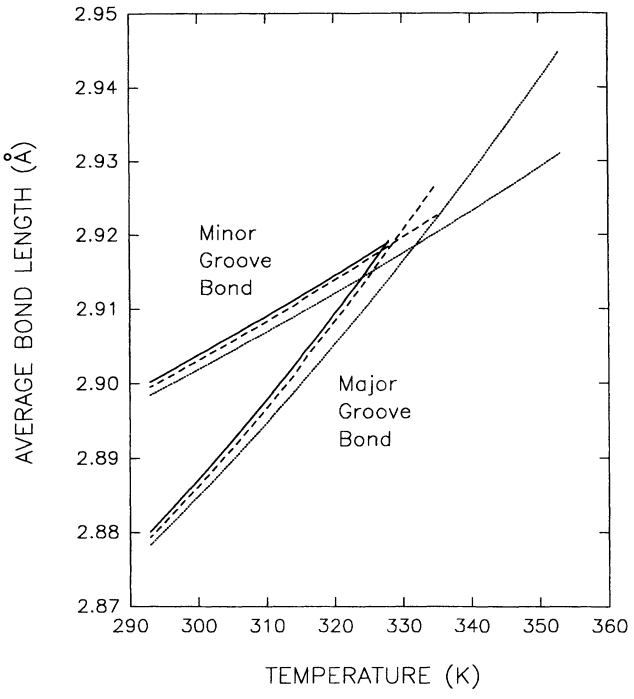

FIG. 4. The average hydrogen bond length is shown as a function of temperature. Solid lines, poly(dA)-poly(dT); dashed lines, the middle A-T base pair of poly(dAAAG)-poly(dCTTT); dotted lines, the A-T base pair of poly(dAGGG)-poly(dCCCT).

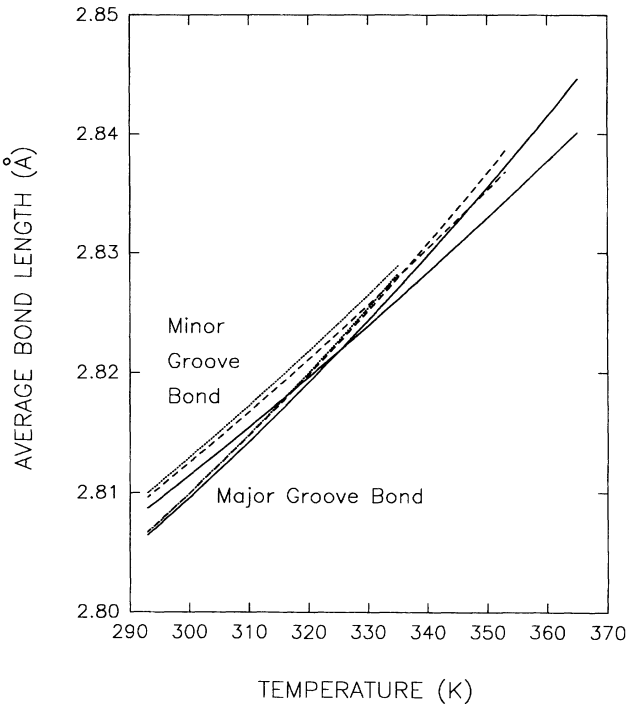

FIG. 5. The average hydrogen bond length is shown as a function of temperature. Solid lines, poly(dG)-poly(dC); dashed lines, the middle G-C base pair of poly(dAGGG)-poly(dCCCT); dotted lines, the G-C base pair of poly(dAAAG)-poly(dCTTT).

the helix. For example, the experimentally determined melting temperatures of poly(dA)-poly(dT), poly(dAT)-poly(dAT), and poly(dAAT)-poly(dATT) are 326.2 K, 318.2 K, and 322.4 K, respectively, in a solution of ionic strength of 19.5 mM Na^{+1}. The melting temperatures of poly(dGAA)-poly(dTTC) and poly(dGTA)-poly(dTAC) differ by 2.3 K in the same environment [1].

Using the MSPA approach, we have calculated the melting temperatures of DNA double helices of same G-C content but different in base-pair arrangements. However, no significant differences in melting temperatures were found among such helices. Melting temperatures calculated for some of the DNA helices are listed in Table II. Detailed examinations on the bond lengths and anharmonic force constants of hydrogen bonds further showed that there is not much difference in the dynamics of these helices. For the DNA helices listed in Table II, the vibrational amplitudes, average bond lengths, and anharmonic force constants as functions of temperature are essentially the same in all those DNA helices with the same G-C content. The small difference from one helix to another is not enough to affect the melting temperature and produce the observed differences. This seems in contrast to a similar calculation by Prabhu, Young, and Prohofsky [6]. In their study, the melting temperatures of several different DNA polymers made with unit cells of two base pairs were calculated and different melting temperatures were observed for DNA helices of same G-C–to–A-T ratio but different in a base-pair sequence. However, they used the instability in the perturbation calculation as a criterion for helix melting, which has been proved inappropriate [14]. Based on the current study, we speculate that the different melting temperatures of DNA helices having the same G-C–to–A-T ratio as shown in Table II might be attributed to water and counter ions surrounding the DNA helices. Because of the difference in structures between helices such as poly(dA)-poly(dT) and poly(dAT)-poly(dAT), one would expect that water molecules and counter ions bind to and interact with such helices differently and thus cause the difference in their melting temperatures. A study on the effects of an enzyme on the dynamics of DNA by Beger and Prohofsky [18] did indicate that ions attached to various locations on poly(dAT)-poly(dAT) produce a larger interstrand breathing motion than poly(dA)-poly(dT). These effects are not taken into account in the present calculation. The different DNA helices in each group of Table II are actually obtained by replacing an A-T base pair with a T-A base pair, or replacing a G-C base pair with a C-G base pair. They are not different in the overall G-C and A-T sequence. Different melting temperatures might be obtained for DNA helices such as poly(dAAAGG)-poly(dCCTTT) and poly(dAGAGA)-poly(dTCTCT) using the MSPA theory due to the reasons mentioned above.

TABLE II. Comparison of calculated melting temperatures with available experimentally observed values and those in Ref. [6].

DNA	G-C (%)	Expt.	Calc.	Ref. [6]
poly(dA)-poly(dT)	0	326.2	328	
poly(dAT)-poly(dAT)	0	318.2	329	318
poly(dAAT)-poly(dATT)	0	322.4	328	
poly(dAG)-poly(dCT)	50	344.5	345	370
poly(dAC)-poly(dGT)	50	350.6	345	366
poly(dAAG)-poly(dCTT)	33.3	337.7	339	
poly(dGTA)-poly(dTAC)	33.3	340.0	339	
poly(dAAC)-poly(dGTT)	33.3	343.4	339	
poly(dGAT)-poly(dATC)	33.3	345.2	339	

[1] W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
[2] J. Marmur and P. Doty, J. Mol. Biol. **5**, 109 (1962).
[3] Y. Gao and E. W. Prohofsky, J. Chem. Phys. **80**, 2242 (1984).
[4] Y. Gao, K. V. Devi-Prasad, and E. W. Prohofsky, J. Chem. Phys. **80**, 6291 (1984).
[5] Y. Kim, K. V. Devi-Prasad, and E. W. Prohofsky, Phys. Rev. B **32**, 5185 (1985).
[6] V. V. Prabhu, L. Young, and E. W. Prohofsky, Phys. Rev. B **39**, 5436 (1989).
[7] Y. Kim and E. W. Prohofsky, Phys. Rev. B **33**, 5676 (1986).
[8] W. Zhuang and E. W. Prohofsky, Phys. Rev. A **41**, 5677 (1990).
[9] Y. Feng, W. Zhuang, and E. W. Prohofsky, Phys. Rev. A **43**, 1049 (1991).
[10] V. V. Prabhu, L. Young, and E. W. Prohofsky, Biopolymers **31**, 355 (1991).
[11] Y. Z. Chen, Y. Feng, and E. W. Prohofsky, Biopolymers **31**, 139 (1991).
[12] Y. Z. Chen, W. Zhuang, and E. W. Prohofsky, Biopolymers **31**, 1273 (1991).
[13] E. W. Prohofsky, in *Biomolecular Stereodynamics IV*, Proceedings of the Fourth Conversation in the Discipline Biomolecular Stereodynamics, Albany, 1985, edited by R. H. Sarma and M. H. Sarma (Adenine, New York, 1985), pp. 21–46.
[14] Y. Feng and E. W. Prohofsky, Phys. Rev. A **43**, 9284 (1991).
[15] N. C. Baird, Int. J. Quantum Chem. Symp. **1**, 49 (1974).
[16] Y. Z. Chen, Y. Feng, and E. W. Prohofsky, Phys. Rev. B **42**, 11 335 (1990).
[17] Y. Feng and E. W. Prohofsky (unpublished).
[18] R. Beger and E. W. Prohofsky, Phys. Rev. B **43**, 5672 (1991).