PHYSICAL REVIEW A VOLUME 43, NUMBER 2 15 JANUARY 1991

Calculation of temperature dependence of interbase breathing motion of a guanine-cytosine DNA double helix with adenine-thymine insert

Y. Feng,* W. Zhuang, and E. W. Prohofsky
Department of Physics, Purdue University, West Lafayette, Indiana 47907

(Received 18 June 1990)

The modified self-consistent phonon approximation is generalized to calculate the critical temperature of a DNA double helix with block inserts of different base pair sequences. An iterative method based entirely on the Green's function is developed to compute efficiently the self-consistent anharmonic force constants and thermal mean-squared vibrational amplitudes in hydrogen bonds. The calculation is carried out for a long guanine-cytosine (G-C) type helix with an insert of d(A)$_4$-d(T)$_4$. The melting-associated behavior is predicted to initiate in major groove bonds in the inserted adenine-thymine (A-T) base pairs at 349 K. This temperature is above the mean-field melting temperature of poly(dA)-poly(dT) due to stabilization of the A—T bonds by the G-C helix and below the mean-field melting temperature of poly(dG)-poly(dC), where dA refers to repeating adenine bases on one strand and dT the repeating thymine bases on the other strand.

I. INTRODUCTION

In an earlier work,[1] we studied the breathing modes in a DNA double helix with adenine-thymine (A-T) type insert. We found that the vibrational amplitude of the average hydrogen bond stretch is enhanced around the inserted A-T block. We suggested that such enhancement in thermal vibrational amplitude could be the physical basis for the low stability at A-T-rich regions in DNA double helices. In natural DNA, local A-T-rich regions are believed less stable and therefore melt at lower temperatures than guanine-cytosine (G-C)-rich regions,[2-4] which leads to initiation of melting from A-T-rich sites. Such open regions offer potential interaction sites. They play important roles in various biological processes involving DNA.

As a continuation of this work, we generalized the modified self-consistent phonon approximation[5] (MSPA) to study the stability of a DNA double helix mediated by "large" defects such as the inserted A-T block in a long G-C-type helix. We report here the estimated melting temperature of a G-C-type DNA double helix with an insert of four A-T base pairs.

MSPA has been used to predict a critical temperature associated with interbase hydrogen bond softening in a DNA double helix. This softening and even apparent breaking of the interbase hydrogen bonds was associated with helix melting and the critical temperature with the strand separation melting temperature. Calculations have been done for a wide range of different sequence DNA homopolymers[6,7] and copolymers[8] as well as defect mediated melting of DNA double helices.[9,10] In the MSPA theory, the interbase hydrogen bonds are modeled by a nonlinear Morse potential.[11] An iterative method is used to compute the mean-squared vibrational amplitude and anharmonic force constant in the hydrogen bonds self-consistently. For DNA homopolymers or copolymers which possess helical symmetry,[12] a set of normal modes of the system can be computed by diagonalizing the dynamic matrix in each stage of calculation, then the mean-square stretch in the hydrogen bonds can be obtained by summing up the contribution from each of these normal modes. In the presence of a defect which breaks the helical symmetry, a Green's-function method[9,10,13] has been developed to compute the mean-square stretch. However, when the degrees of freedom directly affected by the defect are relatively large, this method can be very time-consuming and inefficient. The algorithm we develop here eliminates the calculations of the eigenvalues and eigenfunctions except for generating the initial harmonic Green's functions. All further calculations are entirely based on the Green's-function method. This drastically reduces the computing time and increases the accuracy in numerical calculation. It could be used to efficiently compute the self-consistent solution for a DNA double helix with large defect.

As in Refs. 1 and 14, the effects due to bending and possible conformation change[15] at the A-T block are not taken into account due to the complexity of the problem and will be investigated in further studies.

II. GREEN'S FUNCTION AND MSPA

As described in Ref. 1, the helix is constructed from a perfect double helix, poly(dG)-poly(dC) in the B conformation, and a finite section of double helix with the same conformation but consisting of A-T base pairs. The insertion is done by cutting a section from the perfect G-C helix and replacing it with the A-T block (Fig. 1). The standard B conformation is assumed for the final system. Structural distortion which may occur around the A-T insert are neglected in our calculation due to complexity and will be investigated in further study. Therefore, only the sequence is altered in a standard B conformation. We choose to cut the O(5)—C(5) bond so that the number of valence forces involved is a minimum.

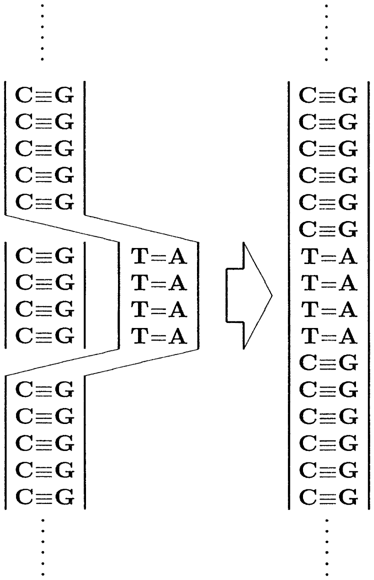

FIG. 1. The construction of the helix is illustrated. The insertion is done by cutting a section from the perfect helix, poly(dG)-poly(dC), and replace it with an oligomer d(A)$_4$-d(T)$_4$.

The DNA double helix can be described by the eigenproblem

$$(\underline{F}-\omega^2\underline{I}+\underline{C}+\underline{\bar{C}})\mathrm{q}=0 \ , \tag{1}$$

where $\underline{F}$ is the matrix of force constants for the helix system we start with, which consists of the homopolymer poly(dG)-poly(dC) and an oligomer d(A)$_4$-d(T)$_4$. The matrix $\underline{C}$ is the change in $\underline{F}$ necessary to bring about the severing of the homopolymer and the reconnecting of the A-T block to the two semi-infinite strands, the matrix $\underline{\bar{C}}$ is the change in $\underline{F}$ due to anharmonicity in hydrogen bonds which are allowed to soften as temperature increases, ω^2 and **q** are the eigenvalues and eigenfunctions.

We are interested in the hydrogen bond stretch force constants, thermal mean-square stretch amplitudes, and other quantities at a given temperature. These quantities may be directly obtained from the normal-mode frequencies and eigenfunctions if the secular equation can be solved. In the current problem, however, such equations cannot be solved for normal modes of the system because of the anharmonic force constants.

In the MSPA theory, an iterative method is introduced to compute the above quantities self-consistently. In this calculation, an initial set of force constants for the hydrogen bond stretch is obtained by fitting to the Lippincott-Schroeder potential[16] for hydrogen bonds. Then for DNA homopolymers or copolymers, the secular equation is solved to obtain the normal-mode frequencies and eigenfunctions from which an anharmonic force constant is calculated for each hydrogen bond based on the Morse potential. These force constants are then used to calculate a new set of normal-mode frequencies and eigenfunctions. The process is repeated until it reaches the self-consistency which is indicated by a convergence in the force constant for each hydrogen bond into a stable value within a common small tolerance.

In the presence of a defect which breaks the helical symmetry, one cannot solve the secular equation for the normal modes by simple diagonalization. However, if the defects are confined within a finite region in the DNA double helix, we can use the Green's-function method to compute the above quantities.

In our approach only the matrix $\underline{\bar{C}}$ depends on temperature. We may define a Green's function for the system at fixed temperature with only structural defect (i.e., $\underline{\bar{C}}=0$),

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

and a Green's function for the complete system,

$$\underline{\bar{G}}=(\omega^2\underline{I}-\underline{F}-\underline{C}-\underline{\bar{C}})^{-1} \ . \tag{3}$$

It can be shown that[17]

$$\underline{\bar{G}}=\underline{G}+\underline{G}\,\underline{\bar{T}}\,\underline{G} \ , \tag{4}$$

where

$$\underline{\bar{T}}=(\underline{\bar{C}}^{-1}-\underline{G})^{-1} \ . \tag{5}$$

The Green's function $\underline{G}$ is temperature independent and needs to be calculated only once from the Green's function of the initial helix system. This can be found from the normal-mode frequencies and eigenfunctions.[1,14,18] Then Eqs. (4) and (5) can be used to compute the Green's function $\underline{\bar{G}}$ of the complete system at any iterative step for a set of trial anharmonic force constants in hydrogen bond stretches. The thermal mean-squared vibrational amplitude for the ith hydrogen bond may be obtained by[14,18]

$$D_i=\frac{\hbar}{\pi}\int d\omega\,\mathrm{Im}[\bar{G}_{ii}(\omega^2)]\coth\left[\frac{\hbar\omega}{2kT}\right] \ . \tag{6}$$

The anharmonic force constant corresponding to this stretch amplitude is then calculated from

$$\phi=\frac{\int du\, e^{-u^2/2D}\dfrac{d^2V}{du^2}}{\int du\, e^{-u^2/2D}} \ , \tag{7}$$

where

$$V(u)=V_0(e^{-2a(\langle R\rangle+u)}-2e^{-a(\langle R\rangle+u)}) \tag{8}$$

is the Morse potential. Here V_0 and a are parameters which characterize the potential, $\langle R\rangle$ is the displacement of the center of the distribution function $\exp(-u^2/2D)$ from the equilibrium position due to thermal expansion. It is assumed to be a function of the mean-squared stretch amplitude. The determination of $\langle R\rangle$ has been given in earlier publications.[6]

The computed force constant ϕ is expected to differ from the initial value but be closer to the true effective anharmonic force constant of the hydrogen bond. We may then use this force constant in Eqs. (4)–(8) iteratively to calculate another force constant which is even closer to the exact anharmonic force constant. The process may be repeated until the force constant converges to a stable value. This stable value is the self-consistent anharmonic force constant.

III. NUMERICAL CALCULATION AND APPROXIMATIONS

As pointed out in the preceding section, self-consistent anharmonic force constants can be calculated at a given temperature by using the standard iterative method which uses the calculated anharmonic force constant of the current iteration as the initial force constant for the next iteration. The method, however, is very slow in reaching the convergence. A large number of iterations is required in order for the relative change in force constant to be less than a given small tolerance. In this calculation, we used a different approach in computing the self-consistent anharmonic force constant, which can improve the speed of convergence appreciably.

Let ϕ_i be an initial guess for the anharmonic force constant for a particular hydrogen bond and ϕ_f be the force constant calculated from Eqs. (4)–(8). Due to the fact that ϕ_f is always closer to the true anharmonic force constant than ϕ_i, the function defined by the difference between ϕ_f and ϕ_i

$$f(\phi_i)=\phi_f(\phi_i)-\phi_i \tag{9}$$

is a monotonic function of ϕ_i and the self-consistent anharmonic force constant is given by $f(\phi_0)=0$. $f(\phi_i)>0$ indicates that the initial guess for the anharmonic force constant is too small, while $f(\phi_i)<0$ indicates the initial guess is too large. Then any root finding algorithm may be used to find the self-consistent force constant ϕ_0. We choose the secant[19] method in our calculation because it does not require initial guesses to bracket the root. The effect of collective motion is neglected in writing Eq. (9). In fact, both ϕ_f and f should be functions of ϕ_i for all of the hydrogen bonds considered self-consistently. However, the effects from changes in force constants of other hydrogen bonds are secondary and Eq. (9) is valid at least in the lowest-order approximation.

Two initial guesses for the anharmonic force constant are required for each hydrogen bond in the secant method. The harmonic force constant obtained from the Lippincott-Schroeder potential is chosen to be the first one. At room temperature, the anharmonic force constant can be determined self-consistently by any reasonable second choice. For example, the calculated force constant ϕ_f from the first choice (harmonic force constant) is taken to be the second guess in our calculation. At higher temperatures, we use an extrapolation scheme to predict the anharmonic force constant from the self-consistently determined force constants at previous temperatures. This predicted force constant is usually very close to the true self-consistent force constant and serves as the second choice in the secant method.

All bonds are treated simultaneously in each step of the iterative calculation. The secant method is only used to predict values to start with. Once we pick these initial values, they are used in Eqs. (5)–(7) to obtain a new set of force constants. The final outcome is independent of the trial values as any self-consistent solution should be. But only a few steps are required to achieve self-consistency. As a criterion for self-consistency, we require that the relative changes from the previous iteration and from the trial values for each hydrogen bond be less than a common small tolerance. This ensures that convergency is achieved for all of the hydrogen bonds considered. This is necessary because converging speeds are different for different hydrogen bonds. The new criterion is more strict than those used in previous self-consistent calculations.[5,7–9]

As discussed in Ref. 14, a set of effective forces is used to replace the long-range interactions across the interfaces, which are too complicated to be treated exactly.

The calculation of the Green's function $\underline{G}$ for the system involving only structural defects has been described in detail in Refs. 1 and 14 for in-band modes. In the current self-consistent calculation, however, the integral in Eq. (6) has to be done repeatedly. To calculate the integral accurately is a challenge because the imaginary part of the Green's function $\underline{\bar{G}}$, which includes the density of states implicitly, is an irregular function of frequency. This is particularly true around the discrete A-T modes which are shifted in frequency as the force constants change. To overcome this difficulty, we assign a relatively large imaginary part to ω^2 for all modes, including local modes. Besides smoothing the density of states as a function of frequency, this also enables us to treat the entire spectrum uniformly as in-band modes. In this calculation, an imaginary part of $\gamma=\omega^2/100\pi$ is introduced to ω^2, while the mesh size of the integral in Eq. (6) is chosen to be $\Delta\omega^2=\gamma/2=\omega^2/200\pi$. The integral is done using the trapezoid rule.

The dispersion branches of poly(dG)-poly(dC), which lie between 10 and 300 cm^{-1} are treated exactly. These include bands 1–31. The high-frequency modes are treated as dispersionless and their contributions to the thermal mean-square amplitudes are assumed to be constants. The contribution from the very low-frequency (below 10 cm^{-1}) vibrations is expected to be very small and is neglected in the calculation.

The hydrogen bonds in the four inserted A-T base pairs and two G-C base pairs next to the A-T block on each side of it are treated self-consistently. This brings the total number of hydrogen bonds treated self-consistently to 20, which specifies the dimension of the matrix $\underline{T}$. Hydrogen bonds in all remaining base pairs are considered harmonic. The Morse potentials for these 20 hydrogen bonds are the same as these for the corresponding hydrogen bonds in the respective homopolymers.

IV. RESULTS AND DISCUSSIONS

We carried out the self-consistent calculation as described above for a long G-C helix with a d(A)$_4$-d(T)$_4$ insert. Starting from room temperature (293 K), we computed the self-consistent anharmonic force constants of the hydrogen bonds stretch and the thermal mean-squared vibrational amplitudes for the 20 hydrogen bonds mentioned in the preceding section. The condition for self-consistency used in this calculation is

$$\left|\frac{\phi_f-\phi_i}{\phi_i}\right|<0.0001, \quad \left|\frac{\phi_f-\phi'_f}{\phi'_f}\right|<0.0001 ,$$

for each of the 20 hydrogen bonds, where ϕ'_f is the calculated force constant at the last iteration.

The softening in the major groove bond is largest for each base pair as is observed in homopolymers and other DNA polymers. Therefore, close attention is paid to the major groove bonds in observing the softening of the helix. In Fig. 2, the self-consistent mean-squared vibrational amplitudes for the major groove bonds are shown as functions of temperature. The amplitudes are all increasing functions of temperature. The two A-T base pairs in the middle have the largest amplitudes and fastest increase as temperature goes up. The amplitudes in the next two A-T base pairs are smaller because they are closer to the G-C base pairs and more stabilized by the G-C base pairs. On the other hand, the amplitudes in the two G-C base pairs next to the A-T block are slightly larger than in other G-C base pairs due to destabilization by the A-T block.

The self-consistent anharmonic force constants for the major groove bonds are shown in Fig. 3. The force constant decreases as temperature increases as a result of bond softening. Here again the stabilization of the A-T base pairs by the G-C helix and the destabilization of G-C base pairs close to the A-T block by the A-T insert are obvious.

In Fig. 4, we show the average potential energy in each of the major groove bonds at different temperatures. The

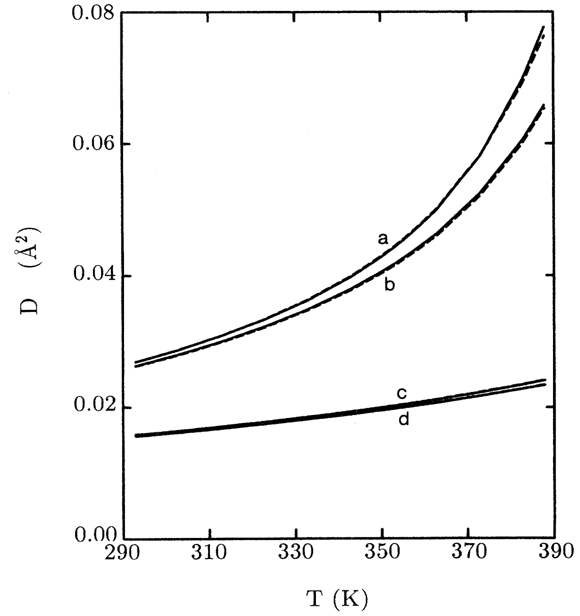

FIG. 2. The thermal mean-squared vibrational amplitudes in the hydrogen bonds near major groove are shown as functions of temperature. The labels indicate (curve *a*) A-T base pairs in the center; (curve *b*) A-T base pairs adjacent to G-C base pairs; (curve *c*) G-C base pairs nearest to the A-T insert; (curve *d*) G-C base pairs next nearest the insert. Since this helix does not have inversion or C_2 symmetry, the behavior of the bonds on either side of a line between the central A-T base pairs is slightly different. The dashed line is for the base pair in the direction of 3′ to 5′ in A and G strands from the center line. The solid line is the base on the other side of the center line.

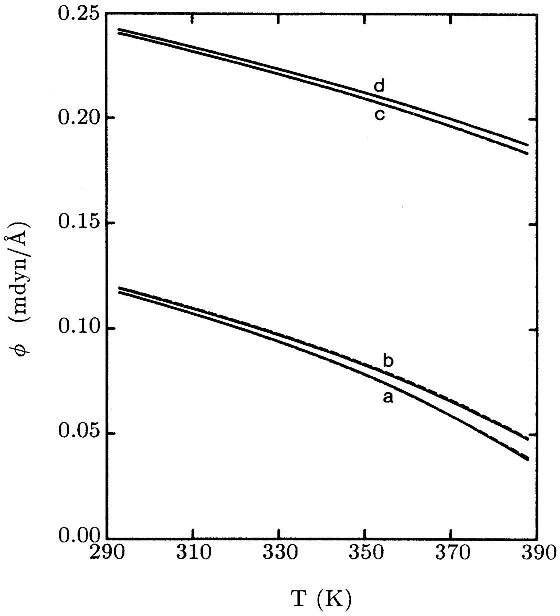

FIG. 3. The self-consistent anharmonic force constants in the hydrogen bonds near the major groove are shown as functions of temperature. The legends are the same as in Fig. 2.

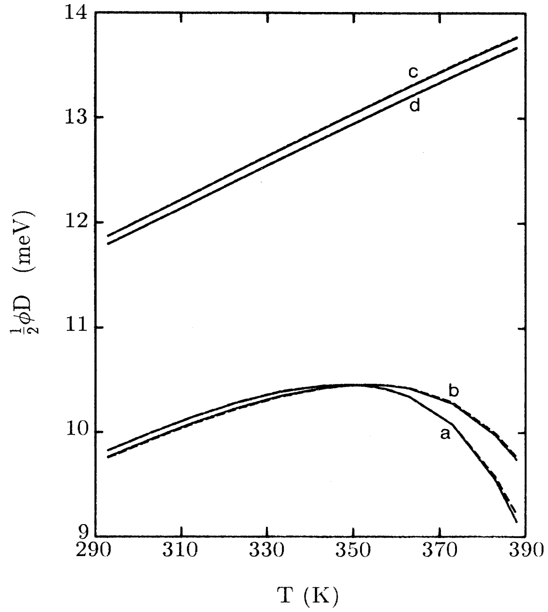

FIG. 4. The average potential energies in the hydrogen bonds near the major groove are shown as functions of temperature. The legends are the same as in Fig. 2.

average potential energies increase with temperature for G-C base pairs in the temperature region considered here, while the average potential energy in the A—T bonds starts to drop after 349 K, which indicates the breakdown of solid behavior in the A-T base pairs. It is this breakdown that is associated with melting of the system and we define this temperature as the lowest turning point in average potential energies.[20,21] The average potential energies in the minor groove bonds and the middle bonds in G-C base pairs are not shown here. They either do not exhibit decrease when temperature increases in the temperature range considered or start decreasing at a temperature higher than 349 K. Therefore we can conclude that melting is likely initiated in the A-T insert region and the major groove bond should be first broken at the melting temperature.

Besides the small softening effect in the G-C base pairs close to the A-T block, nothing extraordinary is observed in the G-C part. The calculation suggests that the induced melting could be local. The presence of a local defect like the A-T insert may have only a small effect on the global behavior of the helix. The nearest G-C base pair could likely melt before bulk G-C but not by too much. Due to the limitations of the current MSPA calculation, which is only expected to provide some indications of strand separation by associated phenomena, we cannot exactly determine when the hydrogen bonds in the G-C base pairs break because the self-consistent solutions at higher temperatures are unphysical.

Melting is initialized in the inserted $d(A)_4$-$d(T)_4$ block around 349 K, which is about 30 K above the mean-field melting temperature for poly(dA)-poly(dT) (Ref. 21) and below the mean-field melting temperature of poly(dG)-poly(dC) which is believed above 400 K using the current model.[20,21] This is qualitatively in agreement with experimental observations. The helix melts at a higher temperature than poly(dA)-poly(dT) due to the stabilization effect by the G-C helix. The G-C base pairs close to the A-T block are destabilized by the A-T block but the effect is not strong enough to induce melting in the G-C base pairs.

The pattern seen here is similar to the results in our most recent mean-field calculations.[20,21] In these calculations, no complete breakdown of self-consistency above a critical temperature is found. Self-consistent solutions are found at high temperatures but they are unphysical and cannot represent a correct solution to what is an effective harmonic system. The decrease in energy in the hydrogen bonds with temperature indicates that no physical solidlike MSPA solution exists. This breakdown of solidlike behavior is assumed to be correlated to the melting of the physical system.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grant No. GM24443 and Office of Naval Research Contract No. N00014-86-K-0252.

---

[*] Present address: Department of Physics, National University of Singapore, Singapore 0511.

[1] Y. Feng and E. W. Prohofsky, Biophys. J. **57**, 547 (1990).

[2] H. C. Spatz and D. M. Crothers, J. Mol. Biol. **42**, 191 (1969).

[3] A. Wada and A. Suyama, in *Biomolecular Stereodynamics IV*, Proceedings of the Fourth Conversation in the Discipline Biomolecular Stereodynamics, Albany, NY, 1985, edited by R. H. Sarma (Adenine, Guilderland, NY, 1986), pp. 21–46.

[4] A. Wada and A. Suyama, Prog. Biophys. Mol. Biol. **47**, 113 (1986).

[5] Y. Gao and E. W. Prohofsky, J. Chem. Phys. **80**, 2242 (1984).

[6] Y. Gao and E. W. Prohofsky, J. Chem. Phys. **80**, 6291 (1984).

[7] Y. Kim, K. V. Devi-Prasad, and E. W. Prohofsky, Phys. Rev. B **32**. 5185 (1985).

[8] V. V. Prabhu, L. Young, and E. W. Prohofsky, Phys. Rev. B **39**, 5436 (1989).

[9] Y. Kim and E. W. Prohofsky, Phys. Rev. B **33**, 5676 (1986).

[10] V. V. Prabhu, L. Young, K. W. Awati, W. Zhuang, and E. W. Prohofsky, Phys. Rev. B **41**, 7839 (1990).

[11] N. C. Baird, Int. J. Quantum Chem. Symp. **1**, 49 (1974).

[12] P. W. Higgs, Proc. R. Soc. London, Ser. A **220**, 472 (1953).

[13] W. Zhuang, Y. Feng, and E. W. Prohofsky (unpublished).

[14] Y. Feng, R. D. Beger, X. Hua, and E. W. Prohofsky, Phys. Rev. A **40**, 4612 (1989).

[15] W. K. Olson, M. H. Sarma, R. H. Sarma, and M. Sundaralingam, in *Structure and Expression,* Vol. 3, *DNA Bending and Curvature,* Proceedings of the Fifth Conversation in the Discipline Biomolecular Stereodynamics, Albany, NY, 1987 (Adenine, Guilderland, NY, 1988).

[16] R. Schroeder and E. R. Lippincott, J. Phys. Chem. **61**, 921 (1957).

[17] H. Böttger, *Principles of the Theory of Lattice Dynamics* (Physik-Verlag, Weinbeim, 1983).

[18] B. F. Putnam, L. L. Van Zandt, E. W. Prohofsky, and W. N. Wei, Biophys. J. **35**, 271 (1981).

[19] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes, The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1986).

[20] Y. Feng and E. W. Prohofsky (unpublished).

[21] Y. Z. Chen, Y. Feng, and E. W. Prohofsky, Phys. Rev. B **42**, 11 335 (1990).