Self-consistent calculation of localized DNA vibrational properties at a double-helix–single-strand junction with anharmonic potential

W. Zhuang, Y. Feng, and E. W. Prohofsky
Department of Physics, Purdue University, West Lafayette, Indiana 47907
(Received 25 August 1989; revised manuscript received 24 January 1990)

We study the dynamics of H-bond motion for a model of a replicating DNA fork. The model contains effects both from the creation of the fork and from self-consistent changes in the H bonds due to the creating of the fork. We use a Morse potential to describe the H-bond interactions. The anharmonic behavior does alter the final H-bond dynamics. The anharmonic aspect increases the H-bond fluctuations by more than a factor of 2 over the harmonic fork results in some frequency regions. We also display the frequency dependence of the H-bond motion and suggest spectral features that can be a signature for the existence of forks in DNA samples.

I. INTRODUCTION

In carrying out its biological function the DNA double helix is often split into single strands over a finite region of its length so that the base-pair message can be efficiently read. For each such open state there are two regions where the transitions are made between H-bonded double-helix and unbonded single strands. These regions are often referred to as the replicating fork or transcribing fork, etc., in the chain. In this paper we study the dynamics of a particularly simple model of such a region for the DNA homopolymer poly(dG)-poly(dC). [The notation poly(dG)-poly(dC) means that one strand contains only guanine (G) bases and the other only cytosine (C) bases.]

An earlier work examined the dynamics of a similar fork.[1] In that work it was found that the H bonds of the base pair of the double helix adjacent to the open region had enhanced H-bond stretch due to being near the open region. Since that work we have developed self-consistent methods[2] that show that enhanced stretch motion by itself causes a decrease in the effective self-consistent force constant for atoms joined by realistic potentials, in particular, for the Morse potential.

Given that two factors: (i) proximity to an open region and (ii) enhanced H-bond stretch by itself-each enhance additional H-bond stretch—we expect a synergistic interaction to occur where the H-bond stretch is further enhanced over the earlier calculation by including modified self-consistent phonon approximation[3] (MSPA) effects. In this paper we include MSPA effects and show how much larger the H-bond stretch is due to the synergistic interaction. We present what we expect to be a much better description of the fork dynamics and show the frequency dependence of the H-bond stretch mean-square amplitude as a crude projection of the possible spectral signature of the fork.

These forked structures exist in nature in very long DNA helices. We want to approximate the long helices as infinite helices and we run into the problem of dealing with an infinite system that has a defect. The defect destroys the helical symmetry usually assumed to hold for lattice calculations of the double helix. We solve this problem by constructing a fork which is made out of an original accumulation of perfect helices that retain helical symmetry. Since the original problem satisfies Bloch's theorem that problem can be solved for infinite-size helices by standard helical lattice methods. We then construct the fork by judiciously cutting and joining parts of these perfect helices. Because the cuts and joints take place in a small region of space, the problem is solvable by Green's-function methods and one can avoid dealing with infinite matrices. The cutting and joining is symbolically illustrated in Fig. 1. The lines and tick marks are cartoons for backbones and bases, respectively. The top line represents a single helix of poly(dC). The second connected strand is the base-paired duplex poly(dG)-poly(dC). The next lower line is the single-helix poly(dG). In all cases the arrows indicate the 3′ to 5′ direction of the various backbones. In diagrams below the top one show (a) how the long helices are severed and (b) how they are spliced to form the desired fork. The atomic structure of double-helix poly(dG)-poly(dC) around the cut is displayed in Fig. 2 in which the dotted line represents the cutting plane at which it is believed that the restriction endonuclease enzymes cleave the DNA chain.[4] The guanine and cytosine bases are denoted respectively by G and C. The structure of single-helix poly(dG) and poly(dC) are identical to those in the double helix, respectively.

In setting up such a calculation one has to first determine the conformation of the various segments, i.e., the double-helical section and the two single-strand sections. The doubled helix has a more stable structure and under physiological conditions can be assumed in a standard *B* conformation. In *B* conformation, the DNA has right-handed screw symmetry with pitch angle 36° and pitch height 3.38 Å. The conformation of the single strands is much more problematic. In fact, it is likely to be found in different conformations depending on a host of other factors. To carry out our calculation we need the single strands to be in some repeating configuration. If a re-

peating symmetry or periodicity exists, Bloch's theorem holds, and the dynamic matrix can be block diagonalized. A band structure can then be calculated even if the strands are infinite in length. We need such solutions to write the initial Green's function. If there is no periodicity the initial dynamic matrix is infinite for an infinite strand with no symmetry reduction to finite blocks. One would then have to deal with short single helices to achieve initial solutions needed for the initial Green's function.

In our calculation we assume that the single strands are periodic and stacked so as to conform to standard *B* conformation, except that they are not coupled to each other. We made this choice for several reasons. First, no other unique conformation stands out as more likely. Second, this choice was the simplest for us to implement as the geometrical factors did not require major recalculation. Third, this choice is essentially the most conservative one considering the dynamics we are exploring. The single strands are most like the double strand and a minimum amount of additional fluctuation in the fork may therefore be expected. This calculation can therefore be considered a study of minimum defect-induced H-bond fluctuation enhanced by the synergistic effects discussed, rather than a realistic prediction of a known structural object that occurs in DNA.

There is a further assumption that the conformations chosen for the double and single strands do not distort further, i.e., the conformation near the fork is not distorted by the presence of the fork. This is clearly an assumption, as it is likely that some further local distortions occur at the fork. We do not have a model of such distortion for the time being and our assumption is again the most conservative one. The change in H-bond motion we calculate is thus likely to be a lower limit to the actual case where local distortions and changes in conformation of the single strands occur.

The DNA fork is regarded as a defected system of the original prefect DNA helices. The "defect" involves both the cutting and joining of related bonds. We shall describe the defect in mathematic detail later. It has been generally known that a local defect in an otherwise perfect periodic lattice can bring about significant changes in the dynamic behavior of the system, especially in the region closely surrounding it. New modes may emerge from band gaps where no vibrations are allowed in the perfect system. These modes decay faster than exponentially with distance from the defect and are called pure localized modes. Also, amplitudes of in-band modes can be either enhanced or suppressed significantly around the defect depending on the nature of the bands. Green's function methods have been known to be very useful in this field because they greatly reduce the size of the problem. In many cases the physical quantity of interest can

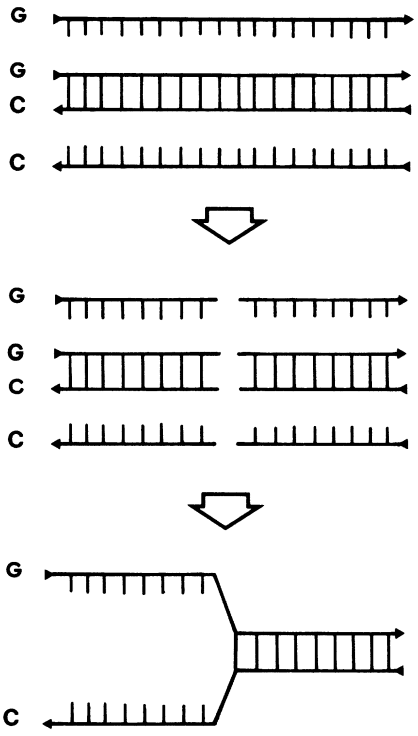

FIG. 1. A DNA fork is constructed from three perfect helices, poly(dG), poly(dG)-poly(dC), and poly(dC) by cutting each of them in half and joining one half of poly(dG) and poly(dC) with that of poly(dG)-poly(dC), respectively. The arrows are along the 3′ to 5′ direction.

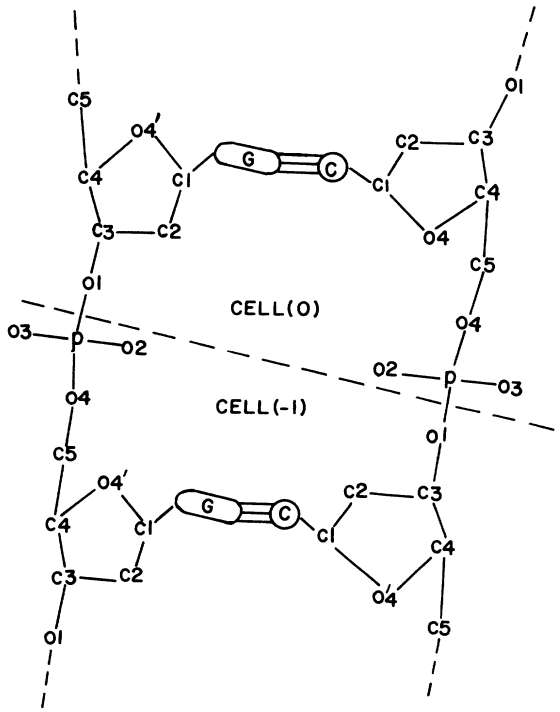

FIG. 2. Portion of two unit cells of poly(dG)-poly(dC). The dashed line indicates the position of the cut. (See Ref. 15.)

be readily expressed by the ensemble average of an appropriate Green's function.[5]

In carrying out the calculations described above we make use of an important property of the Green's-function approach. One can "chain" Green's-function calculations. If two distinct defects are introduced one can introduce them one at a time. For example, the perfect-helix Green's function is generated from the eigenvalues and eigenvectors of the perfect helix. One can generate an intermediate Green's function by introducing the cuts and joints that create the fork. With this intermediate Green's function replacing the perfect Green's function one can now introduce those source terms which arise from self-consistent bond softening and calculate the final Green's function. From the final Green's function one can determine the dynamics of the fork.

The reason that this is particularly useful has to do with the sizes of the various source terms. The initial defect to make a fork involves a large number of altered force constants. This source need be applied once to generate the intermediate Green's function. The self-consistent correction involves only the three H bonds in the cell adjacent the open region, but this source is applied many times as it is an iterative self-consistent solution. The large part is done once and the part needing many iterations is kept small.

II. GREEN'S FUNCTION AND SELF-CONSISTENT PROCEDURE

The general eigenvalue equation to be solved for the DNA fork system is

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

where $\underline{F}$ is the force constant matrix for the perfect system, $\underline{C}$ is the force constant matrix which brings about the defect to the perfect system, ω is the eigenfrequency of the system to be solved, q is the eigenvector corresponding to the eigenfrequency, and $\underline{I}$ is the identity matrix. The dimension of the $\underline{F}$ and $\underline{C}$ matrices is $3N\times 3N$ in Cartesian coordinates, where N is the number of atoms in the system considered. For macromolecules like DNA, N is extremely large. But only a small portion of the $\underline{C}$ matrix will be nonzero if the defect is confined to a small region in space. In the presence case, the defect is confined to the boundary of cell (-1) and cell (0). Using $\underline{g}$ to represent the Green's function for the perfect system and $\underline{G}$ for the DNA for system, they are defined, respectively, as

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

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

where

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

and $\underline{T}=\underline{C}(\underline{I}-\underline{g}\,\underline{C})^{-1}$. Using subscripts $\{a\}$ and $\{b\}$ to represent coordinates directly affected by and not directly affected by the defect, respectively, the $\underline{C}$ matrix can be written as

$$\underline{C}=\begin{bmatrix}\underline{C}_{aa} & \underline{0}\\ \underline{0} & \underline{0}\end{bmatrix}\ , \tag{5}$$

where $\underline{C}_{aa}$ represents the nonzero part. Then the matrix $(\underline{I}-\underline{g}\,\underline{C})$ becomes

$$\underline{I}-\underline{g}\,\underline{C}\begin{bmatrix}\underline{I}-\underline{g}_{aa}\underline{C}_{aa} & \underline{0}\\ -\underline{g}_{ba}\underline{C}_{aa} & \underline{I}\end{bmatrix}\ . \tag{6}$$

Taking the inverse of $(\underline{I}-\underline{g}\,\underline{C})$, we get

$$(\underline{I}-\underline{g}\,\underline{C})^{-1}=\begin{bmatrix}(\underline{I}-\underline{g}_{aa}\underline{C}_{aa})^{-1} & \underline{0}\\ \underline{g}_{ba}\underline{C}_{aa}(\underline{I}-\underline{g}_{aa}\underline{C}_{aa})^{-1} & \underline{I}\end{bmatrix}\ . \tag{7}$$

The $\underline{T}$ matrix becomes

$$\underline{T}=\begin{bmatrix}\underline{C}_{aa}(\underline{I}-\underline{g}_{aa}\underline{C}_{aa})^{-1} & \underline{0}\\ \underline{0} & \underline{0}\end{bmatrix}\ . \tag{8}$$

Therefore the $\underline{T}$ matrix has the same size of nonzero elements as the matrix $\underline{C}$. Splitting up the $\underline{G}$ matrix the same way, we can write $\underline{G}_{bb}$ as

$$\underline{G}_{bb}=\underline{g}_{bb}+\underline{g}_{ba}\underline{T}_{aa}\underline{g}_{ab}\ . \tag{9}$$

Because of symmetry in the $\underline{F}$ matrix, $\underline{g}$ is symmetric and we have $\underline{g}_{ab}=\underline{g}_{ba}$. In internal coordinates the $\underline{C}$ matrix is symmetric and so is the $\underline{T}$ matrix.

The matrix element $\underline{g}$ for the perfect helix is calculated from the eigenvalues $\omega_\lambda(\theta)$ and the eigenvectors $\mathbf{q}^\lambda(\theta)$ of the system by[6]

$$g_{ij}^{mn}(\omega^2)=\frac{1}{\pi}\sum_\lambda\left[\mathrm{P}\left[\int_0^\pi d\theta\frac{\mathrm{Re}\{q_i^\lambda(\theta)[(q_j^\lambda)^*(\theta)]e^{i\theta(m-n)}\}}{\omega^2-\omega_\lambda^2(\theta)}\right]\right]-i\sum_k\left[\frac{\mathrm{Re}\{q_i^\lambda(\theta)[(q_j^\lambda)^*(\theta)]e^{i\theta(m-n)}\}}{|d\omega_k^2(\theta_0)/d\theta|}\right]\ , \tag{10}$$

where i and j are component indices of coordinates, m and n are unit cell indices, λ is branch index of the eigenvalues, θ is phase shift between neighbor unit cells and is equivalent to Bloch wave vector in lattice dynamics, and P stands for the principal part of the integration. In addition, the sum in the imaginary part is over those bands which ω intercepts and θ_0 is the position at which the interception happens. These interceptions are singular in the integral.

It is the imaginary part of Green's function that is related to the displacement correlation tensor D in self-consistent phonon theory. It can be derived that the diagonal elements of the tensor become thermal mean-square displacements in internal stretch coordinates. For

the ith H-bond stretch in a DNA unit cell,

$$D_{ii}^{\text{defect}}=\frac{\hbar}{\pi}\int d\omega \coth(\beta\omega/2)\mathrm{Im}G_{ii}(\omega^2) + \sum_{\lambda}\frac{\hbar}{2\omega_\lambda}q_\lambda^i(q_\lambda^i)^*\coth(\beta\omega_\lambda/2)\ , \tag{11}$$

where the defect Green's-function matrix element G_{ii} is now calculated in internal coordinates. The integral over ω covers the range of frequency of the in-band modes and the sum is over the newly emerged pure localized modes mentioned before. The q_λ^i is a localized eigenvector in internal coordinates too. The frequencies of pure localized modes are determined by

$$\det[\underline{I}-\underline{g}_{aa}(\omega^2)\underline{C}_{aa}]=0\ , \tag{12}$$

and their eigenvectors satisfy

$$(\underline{I}-\underline{g}_{aa}\underline{C}_{aa})\mathbf{q}_\alpha=0 \tag{13}$$

for directly affected coordinates, and

$$\mathbf{q}_b=\underline{g}_{ba}\underline{C}_{aa}\mathbf{q}_a \tag{14}$$

for coordinates not directly affected by the defect.

The solution to this point creates the fork but does not involve the additional effect of H-bond softening. Such an effect is introduced by MSPA. In this approach, an effective Hamiltonian H_E with a set of adjustable force constants $\{\phi_{ij}\}$ (called effective force constants) is established as

$$H_E=-\sum_i\frac{1}{2M_i}\nabla_i^2+\tfrac{1}{2}\sum_{i,j}\tfrac{1}{2}\mathbf{u}_{ij}\cdot\overleftrightarrow{\phi}_{ij}\cdot\mathbf{u}_{ij}\ , \tag{15}$$

where $\mathbf{u}_{ij}$ is relative displacement between atom i and j from their mean positions. By expanding the thermal average of the difference between true and effective Hamiltonians in cumulants and keeping only the first-order term, the minimization of free energy of the system requires the effective force constants to satisfy

$$\phi_{ij}=\langle\nabla\nabla V(\mathbf{R}_{ij}+\mathbf{u}_{ij})\rangle =\frac{\int d\mathbf{u}\, e^{-1/2\mathbf{u}\cdot\mathbf{D}_{ij}^{-1}\cdot\mathbf{u}}\nabla\nabla V(\mathbf{R}_{ij}+\mathbf{u})}{\int d\mathbf{u}\, e^{-1/2\mathbf{u}\cdot\mathbf{D}_{ij}^{-1}\cdot\mathbf{u}}}\ , \tag{16}$$

where $\mathbf{R}_{ij}$ is the vector joining the mean positions of atom i and j, $\mathbf{D}_{ij}=\langle\mathbf{u}_{ij}\mathbf{u}_{ij}\rangle$ is the mean-square displacement correlation tensor, and V is the anharmonic true potential of the system. In the case of a hydrogen bond in DNA, we prefer to use R_T to represent $|\mathbf{R}_{ij}|$. The subscript T indicates that the mean H-bond length is temperature dependent. For each H bond in a DNA unit cell, R_T is set by requiring that

$$V(R_T+\mu_0)=V(R_T-\mu_0)\ , \tag{17}$$

where μ_0 is a calculated amplitude of oscillation of thermal phonons. It is taken to be the width of the Gaussian weight functions

$$\exp\left[-\frac{u^2}{2D}\right]$$

at half maximum, which is $2\sqrt{D\ln 4}$. R_T is determined once the explicit form of true potential V and the value of D for that H bond are known.

Now the self-consistent procedure becomes clear by writing a set of related equations in the following order:

$$D=D(G,T)\ , \tag{18}$$

$$R_T=R_T(D)\ , \tag{19}$$

$$\phi=\phi(R_T,D)\ , \tag{20}$$

$$G=G(\Delta\phi)\ . \tag{21}$$

The procedure begins with finding an appropriate initial value of D by going through the following steps: (a) diagonalizing the secular equation, Eq. (1) with $\underline{C}=0$, for an initial set of force constants to get eigenfrequencies and eigenvectors which fit the available experimental data; (b) calculate the Green's-function matrix $\underline{g}$ for the perfect system with the eigenfrequencies and eigenvectors just obtained; (c) construct the $\underline{C}$ matrix according to the structure of the defect; (d) calculate the $\underline{T}$ matrix which is equal to $\underline{C}(\underline{I}-\underline{g}\,\underline{C})^{-1}$; (e) calculate the Green's function $\underline{G}$ for the system with defect using $\underline{G}=\underline{g}+\underline{g}\,\underline{T}\,\underline{g}$; (f) get values of D from Eq. (11).

After the initial values of D are found, they are used in calculating R_T as well as effective force constants ϕ. The difference between the calculated effective force constants and the initial assumed force constants $\Delta\phi$ are then considered as a defect to the Green's function, which was used in calculating the D's. These $\Delta\phi$'s form a new $\underline{C}$ matrix called $\bar{\underline{C}}$. The new Green's function $\bar{\underline{G}}$, which corresponds to the $\bar{\underline{C}}$ matrix, is then calculated by

$$\bar{\underline{G}}=\underline{G}+\underline{G}\,\bar{\underline{T}}\,\underline{G}\ , \tag{22}$$

where $\bar{\underline{T}}=\bar{\underline{C}}(\underline{I}-\underline{G}\,\bar{\underline{C}})^{-1}$. New D's are calculated with $\bar{\underline{G}}$, which leads to new effective force constants. Such a person is then iterated until convergence is reached.

III. CONSTRUCTING THE $\underline{C}$ MATRIX: SEVERING AND REJOINING THE PERFECT HELICES

To cut the perfect infinite helix into two independent halves one has to turn off all existing interactions across the cutting plane in Fig. 2. Here we assume the plane being the boundary between unit cells (-1) and (0). Interactions across this plane include valence forces (bond stretch, bending, and torsion as well as next-neighbor stretch) on the backbone and nonbonded stacking forces between stacked bases where no valence bonding exists. The latter includes short-ranged van der Waals interactions between nearest-neighbor bases and long-ranged electrostatic Coulomb forces.

The form of potential for the stacking force between atom i in one unit cell and atom j another cell is the following:

$$V_{nb}(r_{ij})=\frac{q_iq_j}{(\epsilon_i\epsilon_j)^{1/2}r_{ij}}-\frac{1.85}{r_{ij}^6}+209.2e^{-3.7r_{ij}}\ , \tag{23}$$

where the first term is the Coulomb potential and the other two terms together are the van der Waals potential.[7] ϵ_i and ϵ_j are empirical dielectric constants whose values change linearly with the distance between atoms.[8] This two-atom potential is summed up for pairs of atoms which are up to nine unit cells apart (one helical turn). Such nonbonded interactions are essential elements in determining characteristics of low-frequency acoustic modes,[9] but have little impact on high-frequency optic modes. The resulting dispersion relations of the above model of nonbonded interactions fitted well with experimental neutron-scattering data.[10]

While the valence forces involve no more than the three atoms, as can be seen from Table I, major concern goes to the nonbonded interactions as they involve a large number of pairs of atoms on both sides of the cutting plane. If we would treat each of the above nonbonded interactions between pairs of atoms as an internal coordinate, the resulting $\underline{C}$ matrix would be extremely large. However, we realize that although the above interactions are numerous, most of them are very small and therefore mainly affect the low-frequency acoustic modes. In these the motion of whole bases rather than the motion of individual atoms is found characteristic. Such gross motions are commonly described as compression, shearing, and tilting. We therefore define a new force constant for each such motion and match its potential to that projected out from the potential used in the perfect-helix calculation. The strength of these interactions decays as the distance between unit cells increases. The interactions between the two unit cells which are immediately adjacent to the cutting plane [i.e., cells (-1) and (0)] are more than an order of magnitude larger than, say, the interaction between cells (-1) and $(+1)$. It is for such a reason that the interactions between cells (-1) and (0) were singled out and are treated explicitly. For the double helix, there are four combinations of such base-base interactions, $G^{-1}G^0$, $C^{-1}C^0$, $G^{-1}C^0$, and $C^{-1}G^0$ between cells (-1) and (0), where G and C here represents the guanine and cytosine bases.

For each generalized motion we define a generalized coordinate which describes it, i.e., compression ΔZ, shearing ΔS,and tilting $\Delta\tau$, where ΔZ is the change in distance along the helix axis between the centers of mass of the two cells, ΔS is the change of distance perpendicular to the helix axis between the two centers of mass, and $\Delta\tau$ is the change in angle between the two normal vectors by which each base plane is defined. Assuming these internal coordinates contribute respectively, $\frac{1}{2}f_c(\Delta Z)^2$, $\frac{1}{2}f_s(\Delta S)^2$, and $\frac{1}{2}f_\tau(\Delta\tau)^2$, to the total potential energy, we define a new force constant that can then be determined by setting the newly defined potential equal to the potential defined in Eq. (23) for that particular motion of atoms.[6] For example, the matching condition for compression is

$$\frac{1}{2}f_c(\Delta z)^2=(\Delta z)^2\sum_{i,j}z_{ij}^2\left|\frac{2q_iq_j}{(\epsilon_i\epsilon_j)^{1/2}r_{ij}^5}-\frac{42\times1.85}{r_{ij}^{10}}+209.2\times3.7^2\frac{\exp(-3.7r_{ij})}{r_{ij}^2}\right|, \tag{24}$$

TABLE I. Valence interactions across the boundary of unit cells (-1) and (0).

Interaction[a]	Type	Force constant[b]
P^{-1}(G)-$O1^0$(G)	Stretch	±3.451
P^{-1}(G)-$O1^0$(G)-$C3^0$(G)	Angle bend	±0.772
$O1^0$(G)-P^{-1}(G)-$O4^{-1}$(G)	Angle bend	±0.620
$O1^0$(G)-P^{-1}(G)-$O2^{-1}$(G)	Angle bend	±0.670
$O1^0$(G)-P^{-1}(G)-$O3^{-1}$(G)	Angle bend	±0.670
P^{-1}(G)-$C3^0$(G)	Nonbonded stretch	±0.150
$O1^0$(G)-$O4^{-1}$(G)	Nonbonded stretch	±0.170
$O1^0$(G)-$O2^{-1}$(G)	Nonbonded stretch	±0.397
$O1^0$(G)-$O3^{-1}$(G)	Nonbonded stretch	±0.397
P^0(C)-$O1^{-1}$(C)	Stretch	±3.451
P^0(C)-$O1^{-1}$(C)-$C3^{-1}$(C)	Angle bend	±0.772
$O1^{-1}$(C)-P^0(C)-$O4^0$(C)	Angle bend	±0.620
$O1^{-1}$(C)-P^0(C)-$O2^0$(C)	Angle bend	±0.670
$O1^{-1}$(C)-P^0(C)-$O3^0$(G)	Angle bend	±0.670
P^0(C)-$C3^{-1}$(C)	Next-neighbor stretch	±0.150
$O1^{-1}$(C)-$O4^0$(C)	Next-neighbor stretch	±0.170
$O1^{-1}$(C)-$O2^0$(C)	Next-neighbor stretch	±0.397
$O1^{-1}$(C)-$O3^0$(C)	Next-neighbor stretch	±0.397

[a]P, O, and C under this column are notations for phosphorus, oxygen, and carbon atoms, as in Fig. 2. The superscript indicates the unit cell number, G and C in parentheses denote the guanine and cytosine strands, respectively.

[b]The force constants are in mdyn/Å.

where $\{i\}$ and $\{j\}$ refer to atoms in cells (-1) and (0), respectively. It can be seen from the equation that f_c is independent of ΔZ. The tilting force constants determined this way turned out to be much smaller than those for compression and shearing and are then neglected in order to save computing time. Since the interactions between farther unit cells are less significant, we treat the sum of them as a whole unit. In addition, since compression falls off more slowly than shearing and tilting ($1/r^3$ versus $1/r^n$ where $n>4$), the major effect of this long-range interaction is just a compressional force. Thus we neglected shearing and tilting between these unit cells and only constructed a generalized compression coordinate by summing up individual compression coordinates for pairs of unit cells up to nine cells away from each other.

Values of nonbonded force constants which are related to generating the defect are displayed in Table II while those of valence bonds are listed in Table I. We have also neglected torsion since it is very small compared to other valence forces. The above procedure does introduce an approximation in the formulation of the problem, but we believe that the error introduced is minor and mostly confined to the lowest frequencies below the region of maximum H-bond contribution. To check if the above simplifications are acceptable we reran the diagonalization program of the perfect helix by replacing the old atom-atom nonbonded potential with that of new force constants. The resulting dispersion curves matched almost perfectly with the old ones on high-frequency bands and fairly well on low-frequency bands, except that they were less smooth. To achieve a perfect cut, we used eigenvalues and eigenvectors from the latter diagonalization as a mismatch between forces in one part of our calculation, and those cut in another part would leave a residue of connections between sections cut.

The $\underline{C}$ matrix is defined as the change of force constants brought about by the defect. It spans both the double- and the single-helical coordinates. It is the position of the terms in this matrix that determines which atoms are affected by a particular interaction. Cutting off an interaction means entering a negative value of force constant in the $\underline{C}$ matrix at a location where it severs atoms that are in the chains to be cut. On the other hand, creating an interaction means entering a positive value at a location in the matrix where it connects atoms that previously had no interaction but are on the parts of the helices that are to be joined. The same absolute value of force constants is used for cutting and joining since the atoms are the same and we assume no change in geometry between the atoms in the original and reformed helices. Single standard G or C replaces the identical G or C in the double helix. Joining them to the double helix introduces the same number of interactions as cutting them. For either a G or C strand, the number of cuts as 12 as can be found out by looking at Fig. 2 as well as Table I. In cutting the double helix there are cross-strand interactions that need to be considered in addition to interactions within each single strand. That brings the number of cuts to 27. Therefore the total number of cuts and joints should be 2×12 (single G) $+2\times12$ (single C) $+27$ (double GC)$=75$. In other words, the number of nonzero elements in the $\underline{C}$ matrix is 75.

IV. SELF-CONSISTENT H-BOND CALCULATION

As stated earlier the creation of the fork did by itself increases H-bond stretch. But weakening of H-bond force constants due to the fork-induced stretch could initiate a further synergistic softening of the H bonds and enhancement of stretch. We now use the Green's function from the fork problem to calculate $\bar{G}$ with self-consistent softening also included.

For this calculation of poly(dG)-poly(dC) in *B* conformation there are three hydrogen bonds in one unit cell: the bond adjacent to the major groove N(4)—H $\cdots$ O(6), the bond in the middle N(1)—H $\cdots$ N(3), and the bond

TABLE II. Nonbond interactions across the boundary of unit cells (-1) and (0).

Interaction[a]	Type	Force constant[b,c]
G^{-1}-G^0	Compression	±0.5040
C^{-1}-C^0	Compression	±0.3805
G^{-1}-C^0	Compression	±0.0381[d]
C^{-1}-G^0	Compression	±0.0082[d]
G^{-1}-G^0	Shearing	±0.2065
C^{-1}-C^0	Shearing	±0.2087
G^{-1}-C^0	Shearing	±0.0235[d]
C^{-1}-G^0	Shearing	±0.0088[d]
$(G)_l$-$(G)_r$	Generalized compression	±0.1740[e]
$(C)_l$-$(C)_r$	Generalized compression	±0.1754[e]
$(GC)_l$-$(GC)_r$	Generalized compression	-0.5246[d]

[a]The superscript indicates the unit cell number, *G* and *C* in parentheses denote the guanine and cytosine strands, respectively, while *G* and *C* without parentheses denote guanine and cytosine bases, respectively.

[b]The force constants are in mydn/Å except that the generalized compressions are in mdyn Å/rad^2.

[c]The force constants are negative for the cutting interaction while positive for the joining interaction.

[d]These interactions exist only in the double helix.

[e]These interactions exist only in the single helices.

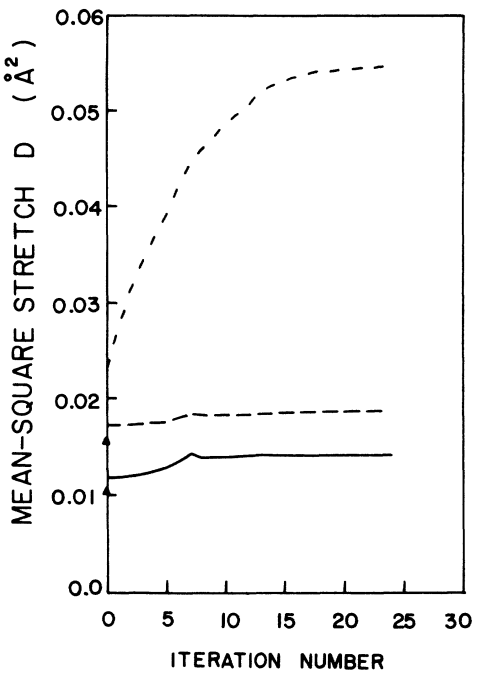

FIG. 3. Mean-square stretch amplitude of H bond near major groove (long-dashed line), near minor groove (short-dashed line), and in the middle (solid line) at each iteration step of self-consistent calculation. The initial values are results in harmonic approximation. The triangles are values for the perfect helix. The two larger amplitudes stretches fall on each other leaving two perfect-helix values.

near minor groove N(2)—H · · · O(2).

We carried out self-consistent calculation for these hydrogen bonds in cell (0), which is the nearest to the double-helix–single-strand junction. All force constants in unit cells above it were not self-consistently adjusted.

Our calculated initial values of D's in unit cell (0) as stated does contain the effect of the fork, i.e., the effect of being adjacent to an open region. They are, however, limited to the harmonic approximation solutions of that problem. They do show enhanced H-bond stretch compared to the initial perfect double helix as shown in Figs. 3–5 and Tables III and IV. This is in general agreement with Putnam and Prohofsky's earlier work on the same system.[1] That is, the hydrogen-bond stretching at the junction is amplified relative to the interior of the double helix. But in that work only two frequency bands were examined and the entire calculation is within harmonic approximation. In present work, we have scanned the entire lower-frequency region 0–230 cm^{-1}, which includes 27 frequency bands from the double-helix poly(dG)-poly(dC), 15 bands from single-helix poly(dG), and 13 bands from single-helix poly(dC). This is the frequency region of large dispersion. Frequency bands higher than 230 cm^{-1} do not change much with Bloch wave vector, therefore they are treated in Einstein model. Our self-consistent calculation showed a further increase in amplification of H-bond stretch vibration at the junc-

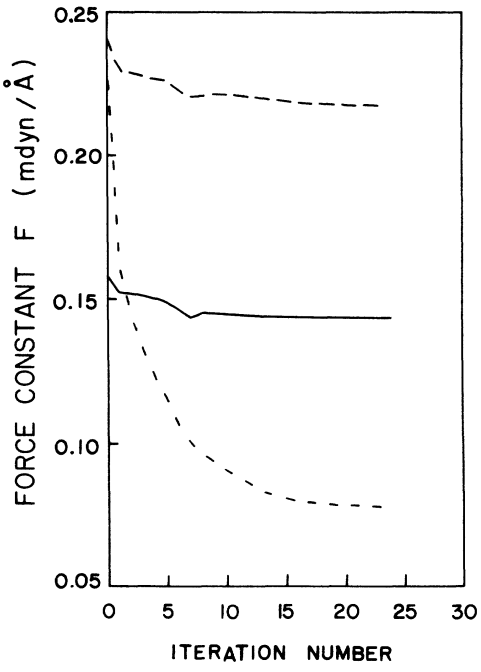

FIG. 4. Effective force constant of H bond near major groove (long-dashed line), near minor groove (short-dashed line), and in the middle (solid line) at each iteration step of self-consistent calculation. The initial values are results in harmonic approximation.

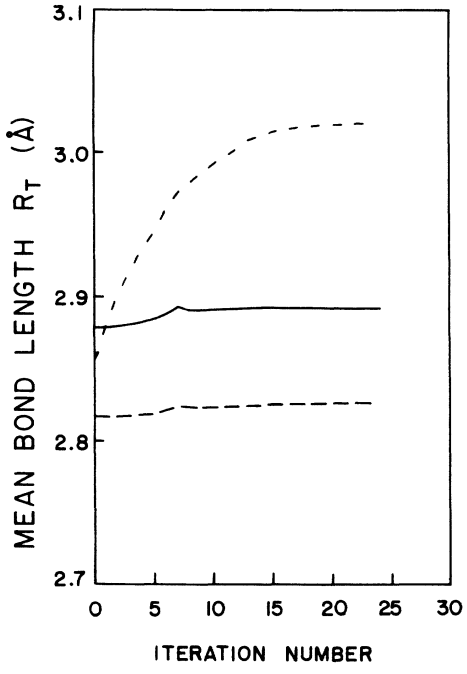

FIG. 5. Mean bond length of H bond near major groove (long-dashed line), near minor groove (short-dashed line), and in the middle (solid line) at each iteration step of self-consistent calculation. The initial values are results in harmonic approximation.

TABLE III. Initial hydrogen-bond force constants and their Morse parameters.

Bond type	Force constant	a (Å^{-1})	V_0 (mdyn/Å)	R_0 (Å)
N(1)—H · · · N(3)	0.158	2.374	0.018 24	2.805
N(4)—H · · · O(6)	0.242	2.881	0.025 72	2.694
N(2)—H · · · O(2)	0.236	2.846	0.025 26	2.698

tion relative to the interior of the double helix compared to that of Putnam and Prohofsky.

The total thermal mean-square displacement fluctuation D, effective force constant F_c, as well as thermal mean bond length R_T of each hydrogen bond at each iteration step are displayed in Figs. 3, 4, and 5, respectively. In all these figures, we have used a solid line to represent the central hydrogen bond N(1)—H · · · N(3), a long-dashed line for the N(4)—H · · · O(6) bond adjacent to the major groove, and a short-dashed line for the N(2)—H · · · O(2) bond near the minor groove.

We see convergence for all three bonds after 24 iteration steps. The convergence criterion has been set for a relative difference between the current force constant and the previous one to be less than 10^{-3} for each hydrogen bond. That is,

$$\left|\frac{\Delta f^{(n)}}{f^{(n)}}\right| = \left|\frac{f^{(n)}-f^{(n-1)}}{f^{(n)}}\right| < 10^{-3} , \tag{25}$$

where n denotes the current iteration step.

We have chosen the Morse potential to serve as the true potential of our DNA hydrogen-bond system as it is found to agree with calculations by Baird.[11] He has calculated bond energy for N—H · · · N as a function of the distance between the heavy atoms by an *ab initio* method. The resulting potential fitted very well to the following Morse potential:

$$V = V_0(1-e^{-\alpha(r-R_0)})^2 - V_0 , \tag{26}$$

where r is the distance between the two nitrogen atoms, $-V_0$ is the minimum value of the potential, R_0 is the position of the minimum, and a has to do with the width of the potential well. V_0, R_0, and a are constant parameters determined from the following observations: (i) x-ray experimental data for positions of atoms in DNA from which hydrogen-bond lengths are known;[12] (ii) hydrogen-bond force constants and dissociation energies;[13] (iii) mean-square displacement D for the perfect DNA helix. Their values in our model are shown in Table II. The anharmonicity of the Morse potential is characterized by its asymmetric shape between the near and far sides of the potential minimum.

For the prefect DNA helix, the mean-square displacement D can be calculated in two ways. One is by summing up band by band contributions from eigenvalues and eigenvectors. The other is by integrating the Green's function over the frequency range covered by the dispersion bands. We do both calculations and use it as a check for our calculated Green's function. The two results agree within 3% of each other.

In Fig. 3, the initial values of D at the beginning of the self-consistent iteration (i.e., $n=0$) are the results from the harmonic approximation. Those at the end of the iterations are the anharmonic results. As can be seen from Table III, the largest increase in total mean-square amplitude D in the harmonic approximation is about 1.5 times larger than that of the perfect helix [the bond near the minor groove, N(2)—H · · · O(2)]. A further factor of 2 increase due to anharmonic effect brings the total increase to 3.5 times. This is equivalent to a thermal fluctuation of 0.23 Å, which is about 8% of the average H-bond length (3 Å) compared to 4% in the perfect helix. It is clear that the possibility of H-bond dissociation is greater around the defect not only due to the presence of the defect but also due to the anharmonicity of H bond. From Fig. 4 we see that the anharmonic effect softens the H-bond force constant near the minor groove by more than 50%.

V. FREQUENCY DEPENDENCE OF H-BOND STRETCH MOTION

In MSPA the total mean-square fluctuation D is calculated. As in Eq. (11) it is determined by the integration of $\mathrm{Im}(\bar{G})\coth(\beta\omega/2)$ over frequencies. We call this integrand thermal fluctuation density $D(\omega)$ since it describes the mean-square stretch fluctuation per unit frequency. One can get a detailed idea about the frequency dependence of thermal fluctuation by plotting it as a function of ω.

TABLE IV. Comparison of mean-square stretches of H bond (in units of Å^2).

Bond type	Perfect helix	Defect helix (harmonic)	Defect helix (anharmonic)
N(1)—H · · · N(3)	0.010 38	0.011 93	0.014 38
N(4)—H · · · O(6)	0.015 74	0.017 39	0.018 94
N(2)—H · · · O(2)	0.015 62	0.023 19	0.054 69

In Fig. 6 $D(\omega)$ are shown for frequencies below 40 cm^{-1} in three cases. They are for the H bond near the (a) minor and (b) major groove and for the (c) middle H bonds, respectively. The dotted lines represent unit cells far away from the junction of the fork and are thus referred to as cell (∞); the dashed lines as well as the solid lines correspond to cell (0), which is the nearest to the junction. The dashed lines are from the harmonic calcu-

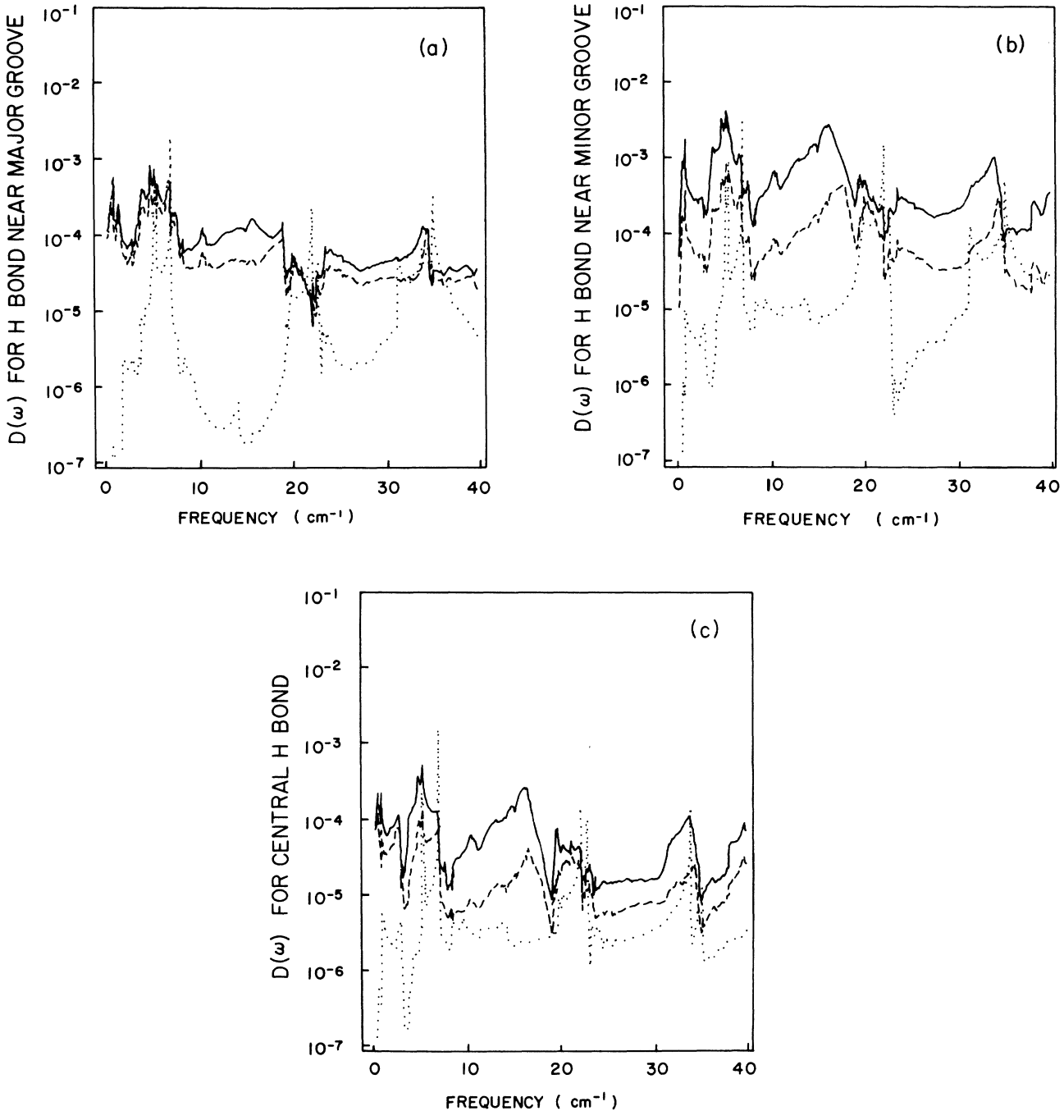

FIG. 6. Mean-square stretch amplitude of H bonds (a) near major groove, (b) near minor groove, and (c) in the middle as a function of frequency between 0–40 cm^{-1}. The dotted lines are for H bonds in cells far away from the junction of the fork, the dashed lines as well as the solid lines represent H bonds nearest to the junction. The dashed lines are from harmonic calculation and the solid lines are from the self-consistent calculation with anharmonic effect. This is the region where the largest increases in stretch amplitude at the junction for all three H bonds with respect to those far away from the junction occur.

lation and the solid ones from self-consistent calculation with anharmonic effect. This is the region where the largest increases (several orders of magnitude) in stretch amplitude in cell (0) occur for all three H bonds with respect to those in cell (∞). This indicates that large coherence H-bond stretch amplitudes have shifted from above 50 cm^{-1} in the perfect DNA double helix to below 40 cm^{-1} near the junction of the DNA fork. This is true in both the harmonic calculation as well as the self-consistent calculation. The effect is greatly enhanced in the self-consistent calculation. The enhancement arises from the fact that in the fork the effective force constants are softened. Again the results here do not take into account the effects of local distortion, possible unstacking, or other conformation changes of the single-helical strands.

The frequency-dependent $D(\omega)$ can be related to a far infrared absorption "signature" for the fork system. An effective dipole moment associated with H-bond displacement x could be written as q^*x, where q^* is an effective charge. The dipole enters into absorption calculations and this factor is proportional to $\langle x \rangle^2$. Since q^* may be assumed independent of frequency, the absorption should be proportional to $D(\omega)$.

The perfect double helix would have much less far infrared absorption. The low-frequency excitations are those involving motion of large sections of the helix. Such modes have long coherence length and ir observation would be limited by k-conservation selection rules.[14] The fork, however, breaks the symmetry of the system and k selection rules no longer hold. The result would be greatly enhanced ir absorption proportional to the $D(\omega)$ factors plotted. Thus our calculation suggests that greatly enhanced far infrared absorption bands are expected for DNA systems with the fork present. But it is an open question whether careful analysis of ir absorption for varying amounts of fork present can lead to detectable absorption that may be assigned to the fork.

ACKNOWLEDGMENTS

This work is supported in part by U.S. Office of Naval Research Grant No. N00014-89-K-0015 and National Institutes of Health Grant No. GM24443.

[1]B. F. Putnam and E. W. Prohofsky, Biopolymers **22**, 1759 (1983).
[2]Y. Gao and E. W. Prohofsky, J. Chem. Phys. **80**, 2242 (1984).
[3]Y. Kim and E. W. Prohofsky, Phys. Rev. B **33**, 5676 (1986).
[4]B.Lewin, *Gene Expression* (Wiley, New York, 1974), Vol. 1.
[5]H. Böttger, *Principles of the Theory of Lattice Dynamics* (Physik-Verlag, Berlin, 1983).
[6]B. F. Putnam, Ph.D. thesis, Purdue University, 1982.
[7]J. M. Eyster and E. W. Prohofsky, Biopolymers **13**, 2005 (1974).
[8]L. Young, V. V. Prabhu, and E. W. Prohofsky, Phys. Rev. A **39**, 3173 (1989).
[9]W. N. Mei, M. Kohli, E. W. Prohofsky, and L. L. Van Zandt, Biopolymers **20**, 833 (1981).
[10]H. Grimm, H. Stiller, C. F. Majkrzak, A. Rupprecht, and U. Dahlborg, Phys. Rev. Lett. **59**, 1780 (1987).
[11]N. C. Baird, Int. J. Quantum Chem. **1**, 49 (1974).
[12]S. Arnott, P. J. Campbell Smith, and R. Chandrasekaran, *Handbook of Biochemistry and Molecular Biology*, 3rd ed. (CRC, Cleveland, 1975), Vol. II, p. 411.
[13]E. R. Lippincott and R. Schroeder, J. Chem. Phys. **61**, 921 (1957).
[14]M. Kohli, W. N. Mei, E. W. Prohofsky, and L. L. Van Zandt, Biopolymers **20**, 853 (1981).
[15]Putnam *et al.*, Biophys. J. **35**, 271 (1981).