Flexibility of Tertiary Structures of Proteins:
Lysozyme and Myoglobin

Motohisa Oobatake and Tatsuo Ooi*

Flexibility in Lysozyme and Myoglobin was examined by simple methods, one calculating displacements of C^α atoms due to small variation of dihedral angle φ (N–C^α bond) and ψ (C^α-C' bond), and the other, first derivatives of total energy by using a tentative repulsive potential between two residues. Since a large displacement of a distance in the negative direction (negative maximum displacement) for an interacting pair, or a large value of first derivative represents occurrence of atomic collision, we can infer the flexibility of each residue. The results show that the flexibility for φ_k is almost identical as for ψ_{k-1} in the opposite direction and terminal regions are more flexible than the rest in the molecule, the range being wider in C-terminal than in N-terminal for both proteins.

INTRODUCTION

A number of three dimensional structures of proteins have been determined by X-ray crystallography[1] during the last decade. Since coordinates of atoms thus determined are not accurate enough to compute an internal energy of a protein, several attempts to obtain an energetically stable conformation, *e.g.,* energy minimization techniques have been applied to proteins.[2-6] Such refinement of coordinates, however, requires much computing time for the calculation of total energy of the protein by summing up all pairwise interactions. The use of coordinates of heavy atoms (C, O, and N), on the other hand, make it possible to examine such properties as flexibility of the stable structure of a protein around the equilibrium, which seems to be related closely to folding and unfolding process, stability of the protein, catalytic reaction mechanism, and allosteric effect, *etc.*

The conformation of a protein may be described as a function of dihedral angles φ (about N-C^α bond) and ψ (about C^α-C' bond) along the main chain, under the assumptions of fixed bond lengths and bond angles, and the standard geometry of a peptide. For the calculation of flexibility about a residue it is necessary to compute energy changes due to variation of φ and ψ from the equilibrium values. Here simple methods for saving the computing time are introduced, considering the nature of interaction energy. First, a change in the distance of a residue pair which is in the interacting range (*e.g.,* within 10 Å) will give informations about the change in energy whether the interaction becomes repulsive or not, because atoms at equilibrium are in an adequate distance near energy minimum favorable for attraction. Computations of such displacements are easy to be done. Second, we shall use a tentative repulsive potential between amino acid residues, instead of calculating all the atomic interactions. These methods will simplify the procedure to estimate flexibility in a protein.

* 大畠玄久, 大井竜夫: Laboratory of Physical Chemistry of Enzyme, Institute for Chemical Research, Kyoto University, Uji, Kyoto.

METHODS

1. Negative Maximum Displacements

Intramolecular energies of a protein molecule are classified into non-bonded, electrostatic, hydrogen bond, torsional, and hydrophobic energy, among which most important one at present is the repulsive part of the non-bonded energy, since atomic collisions do not occur in any conformation. The repulsive part of the non-bonded energy is inversely proportional to 12th power of a pairwise distance, so that the displacement towards a smaller distance of an atomic pair results in a large energy increase when the pair is near potential minimum. We simplify the system by paying attention to the repulsive energy such that movements of every atoms on a residue are replaced to a movement of C^α atom of the residue. That is, changes in energy between residues are estimated by a degree of a displacement of atomic pairs which yields repulsive energy.

According to the above consideration, it is not necessary to compute all the changes in atomic pair distances, since an atomic pair of large separation has no contribution to the stability of the conformation. We have chosen the limit of the distance as 10 Å between C^α atoms, beyond which almost no effect is expected to exist. We adopted conformations, A_f for Lysozyme and K_f for Myoglobin in the previous paper[3] as stable ones which are expressed in terms of dihedral angles $(\varphi_1\psi_1\ldots\ldots\varphi_N\psi_N)$ where N is the residue number. The distance between C^α atoms of i-th and j-th residues, R^s_{ij}, is calculated easily from the set of φ and ψ by using transformation matrices. When a dihedral angle of k-th residue, e.g., φ_k, is rotated by a small amount, namely 1°, mutual distance R^k_{ij} (i,j=1~N) changes from R^s_{ij} by $\varDelta R^k_{ij}$, that is,

$$\varDelta R^k_{ij} = R^k_{ij} - R^s_{ij} \tag{1}$$

Here (ij) pairs were selected on the basis that R^s_{ij} is smaller than 10 Å. $(\varDelta R^k_{ij})_{max}$ is the maximum in absolute value among negative displacements of $\varDelta R^k_{ij}$. The maximum value of the negative displacement $(\varDelta R^k_{ij})_{max}$ indicates a possibility of occurrence in severe atomic collision for the (ij) pair, at a small variation of φ_k or ψ_k. Since the increment can be taken in positive or negative direction, D_k is assigned as the lowest one in absolute value out of $(\varDelta R^k_{ij})_{max}$ in both directions, because a lower value of $(\varDelta R^k_{ij})_{max}$ would correspond to less repulsive collision when compared with the changes of the conformation produced in both directions, e.g., the direction having a lower value of $(\varDelta R^k_{ij})_{max}$ would be favorable than the other direction.

The flexibility occuring in a medium range interaction is estimated by the same procedure, taking pair interactions in the range into consideration, e.g., the interactions involved within ten residues before and behind k-th residue. That is, search of $\varDelta R^k_{ij}$ in Eq. (1) is made over i=1~N and j=i−10~i+10 except for both terminal regions.

2. A Repulsive Energy Function

The repulsive energies between residues are plotted against the separation of the residues from the data in the computations of energies of Lysozyme and Myoglobin.[3] From these plots shown in Fig. 1, an empirical potential as a function of the distance between C^α atoms (as a representative distance between residues) may be assigned in the form of A/r^{10}_{ij} where r_{ij} is the distance between C^α atoms of i-th and j-th residue. The constant A was chosen as 2.8×10^7 Kcal $Å^{10}$/mol to fit the plots in Fig. 1. Then a derivative of
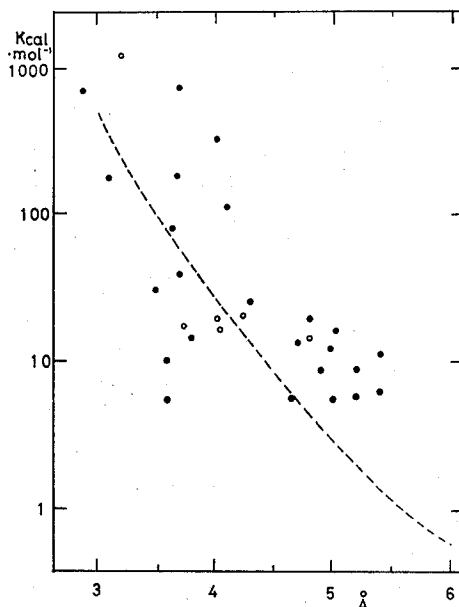
Fig. 1. Plots of nonbonded energy *vs* pair distance between Cᵅ atoms. The dashed line represents a repulsion energy function. The open and closed circles represent the data for Myoglobin and Lysozyme in the previous paper, respectively.

the total energy by the use of the above potential with respect to k-th dihedral angle, φ_k, may be written as

$$G_k = \frac{\partial}{\partial \varphi_k} \left(\sum_{i=1}^{N} \sum_{j=1}^{i} \frac{A}{r^{10}_{ij}} \right)$$

$$= - \sum_{i=1}^{N} \sum_{j=1}^{i} \left(\frac{10A}{r^{11}_{ij}} \frac{\partial r_{ij}}{\partial \varphi_k} \right) \qquad (2)$$

The derivative $\partial r_{ij}/\partial \varphi_k$ can be obtained easily as follows; the deviation of vector r_{ij} is represented by a vector product of $\mathbf{c}_k$, a vector whose components are direction cosines of the bond N_k-C^α_k (for φ_k) and $\mathbf{b}_{kij}$, a vector connecting C^α atom of k-th residue to C^α atom of j-th residue (i>j), or

$$d\mathbf{r}_{ij} = (\mathbf{c}_k \times \mathbf{b}_{kij}) d\varphi_k \qquad (3)$$

Combining Eqs. (2) and (3), we have

$$G_k = - \sum_{i=1}^{N} \sum_{j=1}^{i} \frac{10A}{r^{12}_{ij}} [\mathbf{r}_{ij} \cdot (\mathbf{c}_k \times \mathbf{b}_{kij})] \qquad (4)$$

A change of φ_k from positive direction to negative results only in the change of sign of G_k in Eq. (4).

All the computations were carried out with FACOM 230-60 at the computing center of Kyoto University.

RESULTS

1. Estimate of Flexibility by the Use of Displacement Technique

Computed values of D_k obtained by a small rotation of φ_k and ψ_k plotted against the residue number for Lysozyme and Myoglobin are shown in Figs. 2 and 3, respectively. Distribution of D_k *vs.* the residue number resembles to plots of distances from the center of gravity to every atoms except for N and C terminal regions. It is notable that D_k for φ_k is approximately the same as D_{k-1} for ψ_{k-1}. This is due to the planar structure of a peptide where the bonds $C^{\alpha}_{k-1}-C'_{k-1}$ and $N_k-C^{\alpha}_k$ are almost in parallel with each other. General feature of both patterns of Lysozyme and Myoglobin is that large and small values are repeated alternatively. N and C terminal regions, where D_k values are relatively small, are different in the number of residues, larger for C terminal region than for N-terminal. Since a large value of D_k means possibility of atomic collision, terminal regions
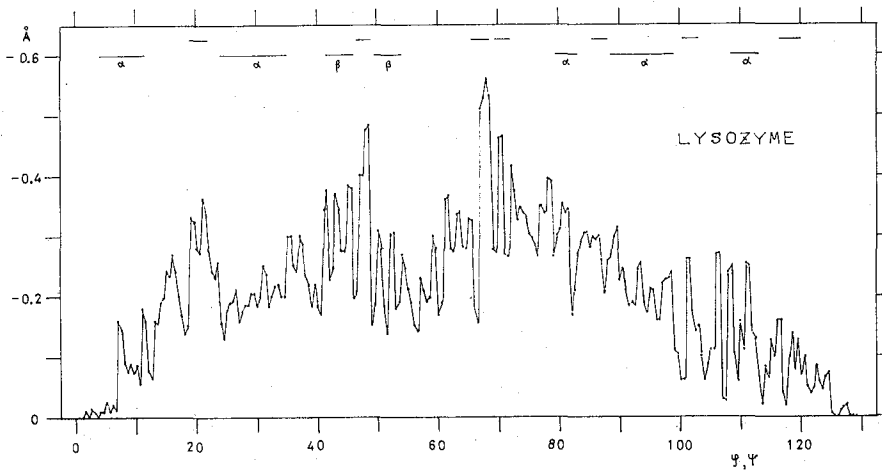
Fig. 2. The negative displacements D_k due to small deviation of dihedral angles of residues ($\varphi_1\psi_1$ ...... $\varphi_N\psi_N$) are aligned in the sequence of Lysozyme. The upper bars shown in the top of the figure represent turn regions, and symbols α and β, α-helix and β-structure regions, respectively.
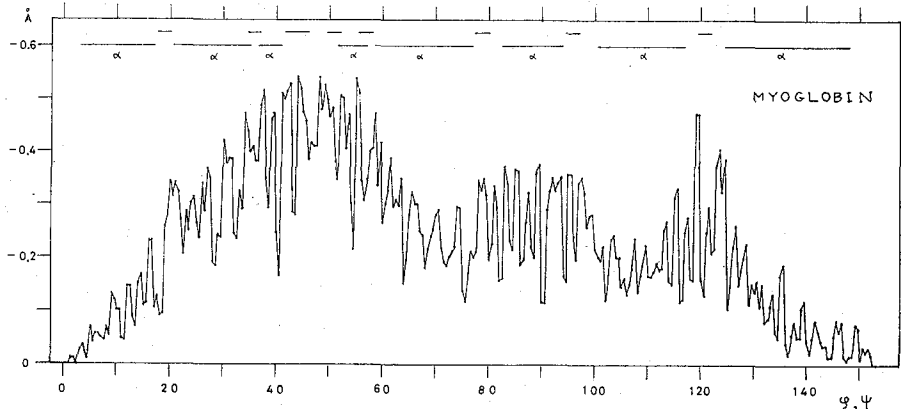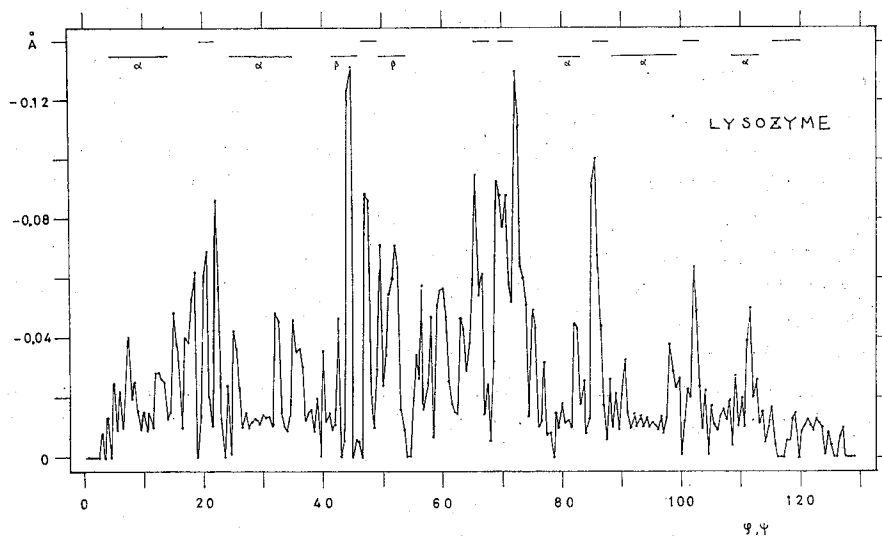

Fig. 3. D_k plots for Myoglobin as in Fig. 2.

Fig. 4. D_k for the short range interaction ten residues before and behind
a residue is plotted in the sequence ($\varphi_1\psi_1$ ...... $\varphi_N\psi_N$) for Lysozyme.
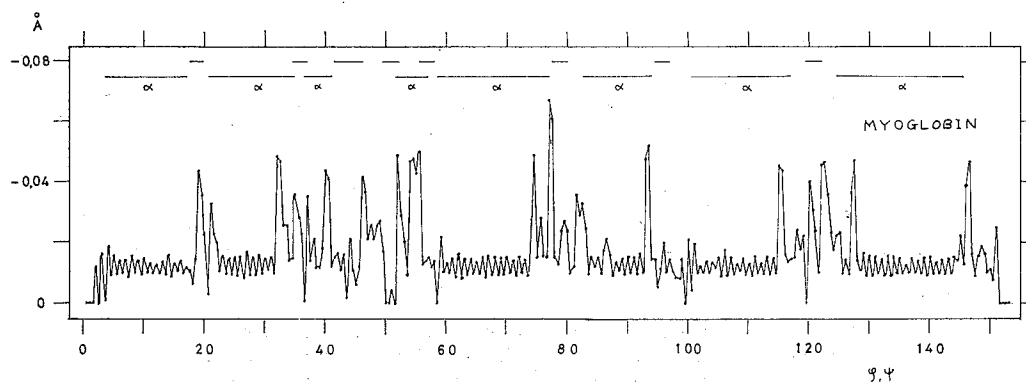

Fig. 5. D_k plots as in Fig. 4 for Myoglobin.

especially near C-terminals are considered to be more flexible than the other regions. The
large values of D_k appeared generally in turn regions, e.g., residue 44 and 68 for Lysozyme
and 44 and 120 for Myoglobin have high D_k values.

Figures 4 and 5 show the pattern of D_k for middle range interaction, taking ten residues
before and behind a residue. The regular zig-zag pattern corresponds to α-helix, which is
demonstrated clearly in Myoglobin. D_k values of turn region and both ends of α-helix
are irregularly high and low.

2. Estimate of Flexibility by the Use of the Repulsive Energy Function

The results described above are based only on extents of negative displacements of
mutual pairs existing within 10 Å, so that the flexibility estimated by this method is quite
rough. On the other hand, the similar results shown in Figs. 6 and 7 are more realistic
for the estimate of flexibility of each residue, since total energy was taken into account,
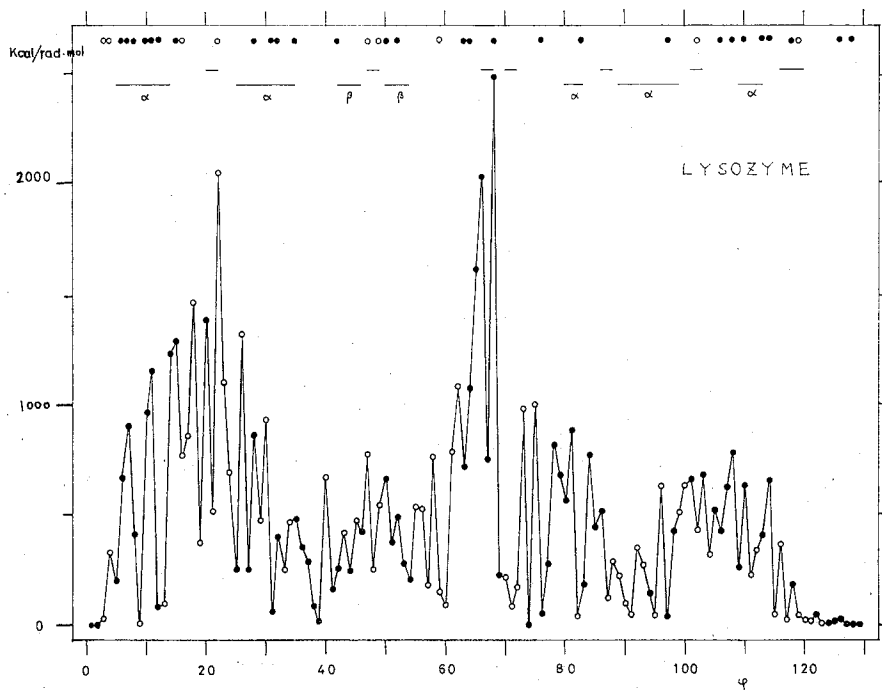though a repulsive part only, for the computations. As shown previously, D_k value for

Fig. 6. First derivatives, G_k, with respect to φ_k are shown in the sequence
of the residue number for Lysozyme. Open and closed circles denote
negative and positive values of G_k. The upper open and closed circles
represent that all the derivatives of interacting pairs are negative and
positive values, respectively.

φ_k is almost the same as for ψ_{k-1}, and, therefore, it is unnecessary to compute a derivative
of the total energy using the present potential for both φ_k and ψ_{k-1}, but only computation
about φ_k may be sufficient.

General pattern of G_k is repetition of high and low values with a period over one to
a few residues. The regions which include derivatives higher than 1000 Kcal/mol/rad are
those from the residue 10 to 27, and from 61 to 68 for Lysozyme, and 60 and 92 to 96 for
Myoglobin. As deduced from the previous results, both terminal regions have low values
of G_k, wider for C-terminal than for N-terminal, and greater in number for Myoglobin than
for Lysozyme, suggesting these regions are more flexible than the other parts of the proteins.

Since fluctuation occurs under thermal motion in the system, the approximate amount
of $\Delta\varphi_k$ for k-th residue would be proportional to RT/G_k. Simple calculation leads to
$\Delta\varphi \sim 0.04°$ for $G_k \sim 800$ kcal/mol/rad and $\Delta\varphi \sim 0.2°$ for $G_k \sim 150$ kcal/mol/rad. In Table I,
the residues which have larger G_k than 800 Kcal/mol/rad and those smaller than 150
kcal/mol/rad were shown in the primary structure of Lysozyme and Myoglobin, respective-
ly. It is clearly shown that fluctuation near C-terminal regions would be large. The more
exact amount of fluctuation, $\langle\Delta\varphi_k^2\rangle^{1/2}$, may be obtained from second derivatives.

Another aspect of the results obtained here is information about directions of $\Delta\varphi_k$.
Since a derivative $G_k = \partial E/\partial\varphi_k$ consists of components, $\partial E_{ij}/\partial\varphi_k$, we can examine those
components for every residues, which are all negative, or negative and positive, or all
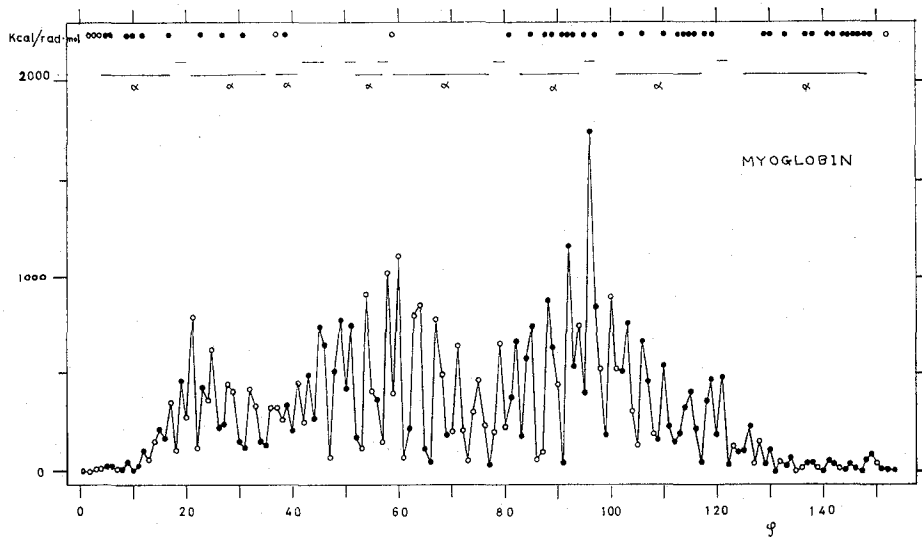positive. These signes of direction are demonstrated in top columns of Figs. 6 and 7,

(461)

Fig. 7. G_k plots as in Fig. 6 for Myoglobin.

Table I. Listings of Flexibility in the Sequences.

Lysozyme

		5				10				15				20				25				30
1	K V F G R C E L A A A M K R H G L D N Y R G Y S L G N W V C
31	A A K F E S N F N T Q A T N R N T D G S T D I G I L Q I N S
61	R W W C N D G R T P G S R N L C N I P C S A L L S S D I T A
91	S V N C A K K I V S D G D G M N A W V A W R N R C K G T D V
121	Q A W I R G C R L

Myoglobin

		5				10				15				20				25				30
1	V L S E G E W Q L V L H V W A K V E A D V A G H G Q D I L I
31	R L F K S H P E T L E K F D R F K H L K T E A E M K A S E D
61	L K K H G V T V L T A L G A I L K K K G H H E A E L K P L A
91	Q S H A T K H K I P I K Y L E F I S E A I I H V L H S R H P
121	G N F G A D A Q G A M N K A L E L F R K D I A A K Y K E L G
151	Y Q G

Residues which have G_k values less than 150 kcal/mol/rad are underlined, and those greater than 800 kcal/mol/rad are shown by dots.

filled circle for positive direction, open circle for negative and not indicated for both direction.   That is, movements of the residues are limited in a certain direction depending on the sign.   Interestingly only a few residues have all negative signs in Myoglobin.

DISCUSSION

The methods described here are convenient for rough estimation of flexibility in a protein without computing total energy precisely. The results, however, might be overestimated, since the methods are taken only displacements of C^α atoms or derivatives of repulsive potentials into consideration. Nevertheless, the results especially by the use of the repulsive potential would indicate the characteristic feature of flexibility, e.g., more freedom around C-terminal regions than the rest in the molecule.

To evaluate fluctuation in a protein of native state, we have to know the exact location of every atoms and then to compute second derivatives with respect to variables, yielding the degree of correlation $\langle \Delta\varphi_i \Delta\varphi_j \rangle$ including the amount of fluctuation $\langle \Delta\varphi_i^2 \rangle$. Performance of such computations requires much computing time and, therefore, the present procedures would be worthwhile to provide rough informations about the flexibility around each residue.

Suppose unfolding of a protein occurs according to a change in environments, then the passway to unfolding may be along gentlest ascend of the energy surface, or movements of dihedral angles would be inversely proportional to G_k. In other words, first step of unfolding would be relatively large rotations about residues which have low G_k values shown in Figs. 6 and 7. Furthermore, it is desirable that components of G_k have the same sign so that the sense of rotation would be favorable for all the components. The residues which satisfy such conditions are quite few in the middle of the sequences, and locate mainly in the terminal regions. Presumably those residues are candidates of initial movements at unfolding.

REFERENCES

(1) "Cold Spr. Harb. Symp. Quant. Biol". University of Tokyo Press, Vol. 36, 1971.
(2) K. Nishikawa, T. Ooi ,Y. Isogai, and N. Saito, *J. Phys. Soc. Japan*, **32**, 1331 (1972).
(3) K. Nishikawa and T. Ooi, *ibid.*, **32**, 1338 (1972).
(4) M. Levitt and S. Lifson, *J. Mol. Biol.*, **46**, 269 (1969).
(5) P. K. Warme, N. Go, and H. A. Scheraga, *J. Computational Phys.*, **9**, 303 (1972).
(6) P. K. Warme and H. A. Scheraga, *ibid.*, **12**, 49 (1973).
(7) K. Nishikawa, M. Oobatake, and T. Ooi, *Bull. Inst. Chem. Res., Kyoto Univ.*, **48**, 102 (1970).