ELSEVIER

Biophysical Chemistry 52 (1994) 107–114

Biophysical Chemistry

Comparison of normal mode analyses on a small globular protein in dihedral angle space and Cartesian coordinate space

Akio Kitao, Steven Hayward, Nobuhiro Go *

Department of Chemistry, Faculty of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606, Japan

Received 21 February 1994; accepted in revised form 26 April 1994

Abstract

Normal mode analyses on the protein, bovine pancreatic trypsin inhibitor, in dihedral angle space and Cartesian coordinate space are compared. In Cartesian coordinate space it is found that modes of frequencies lower than 30 cm^{-1} contribute 80% of the total mean-square fluctuation and are represented almost completely by motions in the dihedral angles. Bond angle and length fluctuations dominate in modes above 200 cm^{-1}, but contribute less than 2% to the total mean-square fluctuation. In the low-frequency modes a good correspondence between patterns of atomic displacements was found, but on average the root-mean-square fluctuations of the Cartesian coordinate modes are 13% greater than their dihedral angle counterparts. The main effect of fluctuations in the bond angles and lengths, therefore, is to allow the dihedral angles to become more flexible. As the important subspaces determined from the two methods overlap considerably, dihedral angle space analysis can be applied to proteins too large for Cartesian coordinate space analysis

Keywords: Normal mode analysis; Dihedral angle space; Cartesian coordinate space; Bovine pancreatic trypsin inhibitor; Important conformational subspace

1. Introduction

Recent studies using the method of principal component analysis applied to molecular dynamics trajectories of proteins in order to determine their collective motions, have demonstrated that fluctuations of only a small number characteristic collective variables, termed principal components, are necessary to account for a high proportion of a protein's motion [1–3]. Principal component analysis is directly related to the normal mode analysis, becoming equivalent in the case of a perfectly harmonic potential energy surface. Although it is known that proteins have anharmonic energy surfaces of a multiple-minima nature, it has been shown that the normal mode analysis can even give information on these anharmonic aspects of the motion [1,4,5]. For example, it has been shown for the polypeptide melit-

* Corresponding author.

SSDI 0301-4622(94)00070-Z

tin, and more recently for the small globular protein, bovine pancreatic trypsin inhibitor, BPTI, that the low-frequency normal modes define a subspace within which a significant proportion of the characteristic, anharmonic motions observed in a molecular dynamics, occur [1,5]. Therefore, normal mode analysis still remains to be one of the most powerful tools available for determining and characterising a protein's dynamics. In this paper we present a comparison of two normal mode analyses on BPTI, one performed in Cartesian coordinates space, CCS, the other in dihedral angle space, DAS. The aim of this comparison is to determine the accuracy of the DAS analysis, where all bond angles and bond lengths are kept fixed. Such a comparison has been previously performed on the polypeptide deca-alanine [6]. However, as indicated above, it is now known that for a protein, motions are dominated by the low-frequency, soft, modes. In the case of deca-alanine, which has a rigid alpha-helix structure, there are relatively few low-frequency normal modes. As it is precisely in these low-frequency modes that one expects the motion to involve almost exclusively the dihedral angles, the possibility arises that the DAS normal mode analysis could be particularly accurate for a protein.

The real point of such a comparison is that the DAS normal mode analysis can be applied to proteins that are too large for practicable application of the CCS normal mode analysis. For example, in the case of BPTI, the CCS analysis requires one to diagonalize a 2658 by 2658 matrix. In the DAS analysis, however, the problem is reduced to the simultaneous diagonalization of two 325 by 325 matrices. This reduction in the size of the problem also gives considerable savings in the energy minimization process. On average the ratio of the number of dihedral angles to Cartesian coordinate variables is about one to eight. However, the DAS analysis has the practical disadvantage of being technically more involved to implement (see Section 2). In this paper we show that despite the approximation of fixing bond angles and lengths, the DAS normal mode analysis provides a way of determining important collective motions of proteins that are too large for a viable CCS analysis.

2. Methods

2.1. Molecule studied

The molecule studied is the small 58 residue globular protein, bovine pancreatic trypsin inhibitor, BPTI. The total number of atoms, N, is 886. The number, N_D, of dihedral angles is 325.

2.2. Conformational energy function

For the DAS calculation the ECEPP energy function was used [7]. For the CCS calculation, the Discover consistent valence force field, CVFF, energy function was used with cross terms using the harmonic potential for the bond lengths [8]. In both normal mode analyses all residues were neutral, the dielectric constant was set equal to 2.0, and no cutoff was applied to the non-bonded interactions. Technical difficulties prevented us from using the same parameter set for both calculations (see below for discussion of this point). The software for the CCS calculation was Discover, and for the DAS calculation, the in-house software, INSPIDAS, was used.

2.3. Energy minimization

The minimized energy conformations were determined by minimizing the energy functions starting at the X-ray structure [9] in the case of the CCS calculation, and in the DAS case starting from the regularized X-ray conformation [10].

2.4. Normal mode analysis

In Cartesian coordinate space

In CCS the normal mode analysis is rather straightforward. After minimization the Hessian matrix is calculated as a function of the mass weighted Cartesian coordinates (Cartesian coordinate multiplied by square root of mass). In our case the Hessian is a 2658 by 2658 matrix. Diagonalization of this matrix yields the eigenvalues, which are the angular frequencies squared, and the eigenvector matrix **U**, the columns of which give the normal mode vectors. Six of the eigenvalues are zero and correspond to the six variables

required to characterize the external motion of the whole protein. This leaves $3N-6$ non-zero eigenvalues.

In dihedral angle space

In DAS this procedure is complicated by the fact that the kinetic energy term cannot be written down as a simple function of the dihedral angle variables, as it can in the case of the mass weighted Cartesian coordinate variables. The internal kinetic energy is calculated as a function of the dihedral angle variables by defining a coordinate system that is attached to the protein by six conditions, three that tie the origin to the centre of mass, and a further three that are collectively known as the Eckart condition which fix the direction of the axes within the body of the protein [11,12]. The normal mode analysis proceeds by calculating the so-called **K** matrix defined as:

$$K_{ij} = \frac{\partial q_i}{\partial \phi_j} \tag{1}$$

in this coordinate system, where q_i is the ith mass weighted Cartesian coordinate, ϕ_j is the jth dihedral angle, and the derivative is calculated at the energy minimum. Knowing this matrix it is possible to write down the kinetic energy in terms of the dihedral angles, by way of the so-called **H** matrix:

$$T = \tfrac{1}{2} \sum_{i,j=1}^{N_D} H_{ij}\dot{\phi}_i\dot{\phi}_j, \tag{2}$$

where

$$H_{ij} = \sum_{k=1}^{3N} \frac{\partial q_k}{\partial \phi_i}\frac{\partial q_k}{\partial \phi_j}. \tag{3}$$

The normal mode analysis is performed by simultaneously diagonalizing the **H** matrix to the identity matrix, and the Hessian to the eigenvalue matrix, yielding an eigenvector matrix **V**. In this case the Hessian and **H** matrices are only 325 by 325 in size. This is a massive reduction in the size of the problem from the CCS case where the matrix is 2658 by 2658.

2.5. Comparing results from the two methods

In the DAS, the dihedral angles in the jth normal mode undergo small amplitude displacements v_{ij} given by the jth column vector of the matrix **V**. The displacement w_{kj} in the kth mass weighted Cartesian coordinate caused by small dihedral angle displacements of the jth DAS normal mode is given by:

$$w_{kj} = \sum_{i=1}^{N_D} K_{ki}v_{ij}. \tag{4}$$

Note there are only N_D $\boldsymbol{w}_j$ vectors. The diagonalization of **H** to the identity matrix means that the w_{kj} satisfy the orthonormality condition:

$$\sum_{k=1}^{3N} w_{ki}w_{kj} = \delta_{ij}. \tag{5}$$

To quantify the similarity of the two normal mode vector sets, $\boldsymbol{u}_j$ from the CCS calculation, and $\boldsymbol{w}_j$, from the DAS, the two minumum energy conformations are brought into their best fit positions reorientating the vector sets $\boldsymbol{w}_j$ and $\boldsymbol{u}_j$ accordingly. The degree of similarity between two particular vectors from the two calculations can be quantified by taking their inner product:

$$c_{ij} = \boldsymbol{w}_i \cdot \boldsymbol{u}_j. \tag{6}$$

c_{ij} gives the value of $\boldsymbol{w}_i$ projected onto $\boldsymbol{u}_j$ or vice versa. If the two vectors are identical, then c_{ij} will equal 1 due to the normality condition.

The space defined by the N_D $\boldsymbol{w}_j$ vectors lies completely within the space defined by the $3N-6$ $\boldsymbol{u}_j$ eigenvectors because all possible patterns of displacements that do not represent overall translations or rotations of the protein, can be represented by linear combinations of the $\boldsymbol{u}_j$ eigenvectors. In other words the $\boldsymbol{u}_j$'s form a complete set. For this reason any $\boldsymbol{w}_i$ can be expressed as a linear combination of the former as:

$$\boldsymbol{w}_i = \sum_{j=1}^{3N-6} c_{ij}\boldsymbol{u}_j, \tag{7}$$

where the c_{ij}'s are given by Eq. (6). It follows from Eq. (7) that:

$$\sum_{j=1}^{3N-6} c_{ij}^2 = 1 \tag{8}$$

as both vector sets satisfy the orthonormality condition. The opposite case of trying to express the CCS eigenvectors $\boldsymbol{u}_j$ as linear combinations of the DAS vectors $\boldsymbol{w}_i$ gives a different result, as the

w_j's do not form a complete set. Physically this is because displacements caused by angle bending and bond stretching are not represented by the w_j's. The quantity,

$$P_j = \sum_{i=1}^{N_D} c_{ij}^2, \tag{9}$$

where now the sum is over the projections of the w_i vectors onto a single u_j vector, measures to what degree the vector u_j is represented in the DAS subspace. If it equals 1 then, the motions are totally representable by dihedral angle motions. If it is less than 1, then it indicates the extent of the contribution from bond length stretching and bond angle bending motions to the atomic displacements.

2.6. Atomic mean-square fluctuation

The atomic mean-square fluctuation derived from the CCS analysis is given by the following expression:

$$\langle x_i^2 \rangle = \frac{k_B T}{m_i} \sum_{j=1}^{3N-6} \left(\frac{u_{ij}}{\omega_j} \right)^2, \tag{10}$$

where x_i is the ith atomic coordinate, m_i, its mass, and ω_j is the angular frequency of the jth CCS normal mode. In the DAS case it is given by:

$$\langle x_i^2 \rangle = \frac{k_B T}{m_i} \sum_{j=1}^{N_D} \left(\frac{w_{ij}}{\omega_j} \right)^2, \tag{11}$$

where in this case, ω_j is the angular frequency of the jth DAS normal mode. The values for the root mean-square fluctuations (RMSF) of the C^α atoms are calculated using these expressions as another way of comparing the two methods. In this work the temperature T was set equal to 300 K.

3. Results and discussion

3.1. Minimized energy conformations

Fig. 1 shows the best fit between the DAS energy minimized structure, and the CCS energy minimized structure. The root-mean-square deviation (RMSD), calculated over main-chain heavy atoms, between these two structures is 0.92 Å.

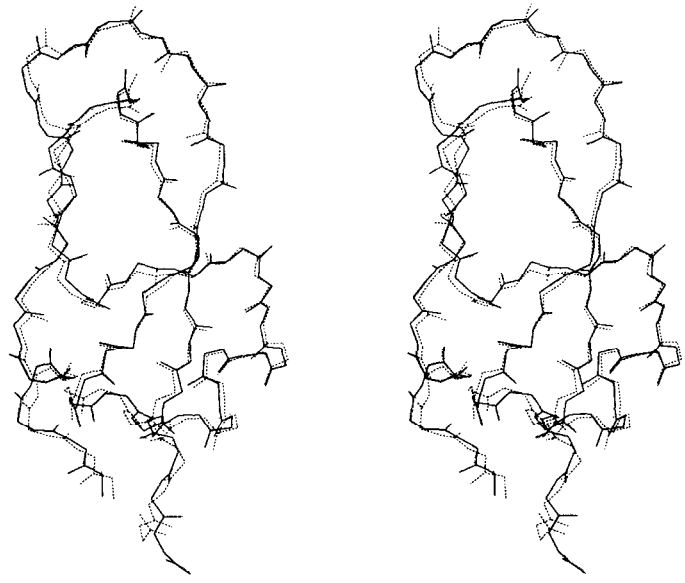

Fig. 1. Stereo view of the X-ray conformation (· · ·), the minimum energy conformation in the CCS (———) and the minimum energy conformation in the DAS (- - - - -).

This small value means that the inner products taken between the normal mode vectors from the two analyses will be largely unaffected by the difference the two minimum energy conformations. The RMSD values after best-fitting to the original X-ray structure are 0.14 Å for the CCS minumized structure, and 0.9 Å for the DAS minimized structure.

3.2. Frequency spectra and RMSF values for the alpha-carbon atoms

The frequency ranges from the two methods, were 5.3–3673.8 cm^{-1} for the CCS analysis, and 6.1–878.7 cm^{-1} for the DAS analysis. Fig. 2 shows the frequency density plotted against frequency from the DAS normal mode analysis, and two CCS normal mode analyses, one using the CVFF parameter set, the other using the OPLS parameter set [13]. We include the OPLS parameter set results only to show that no great difference is seen in using the two parameter sets. A thorough study by Teeter and Case [14] also found that normal mode analysis is rather insensitive to the parameter set used. The OPLS results were used previously in a study comparing the results of a normal mode analysis on BPTI and molecular dynamics simulation [5]. Note that from here onwards, all other CCS results are

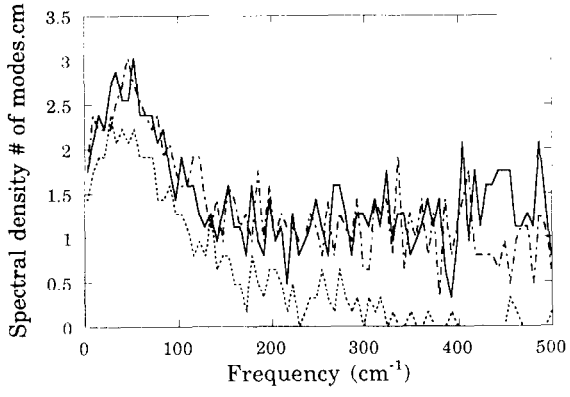

Fig. 2. Frequency distributions obtained from normal mode analyses. The lines (———), (- - - - -) and (-·-·-) represent the results from the CCS analysis using the CVFF energy function, the DAS analysis, and the CCS analysis using the OPLS energy function, respectively.

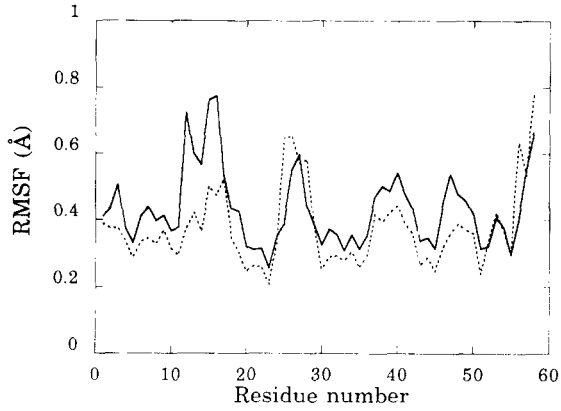

Fig. 3. RMSF (root-mean-square fluctuation) of C^{α} atoms at 300 K. The lines (———) and (- - - - -) represent the results from the CCS and DAS analyses, respectively.

derived from the normal mode analysis using the CVFF energy function. The DAS result shows a consistently lower frequency density. This may not be surprising as the integral of this curve gives the total number of modes: 2652 in the CCS case, compared with 325 for the DAS. Significant differences between the CCS spectra and the DAS spectrum are seen above roughly 200 cm^{-1}, indicating that above this frequency angle bending and bond stretching motions begin to make the major contribution.

Fig. 3 shows the root-mean-square fluctuation (RMSF) values of the C^{α} atoms by the two methods using Eqs. (10) and (11). Overall the peak and trough positions correspond well. Although in the DAS analysis contributions from bond angle bending and bond length stretching are totally neglected, the average RMSF value of the C^{α} atoms from the DAS is 87% of that from the CCS analysis. A very similar picture to that of Fig. 3 is seen in the comparison of side chain atom RMSF values.

The root of the total (including all atoms) mass weighted mean-square fluctuation (normalized by dividing by the total mass) [1] was 0.54 Å from the DAS analysis, and 0.61 Å from the CCS analysis. The fluctuations, therefore, in the CCS case are

[1] This quantity will be referred to as the total fluctuation of the protein from here onwards.

on average 13% greater than those found from the DAS analysis. In the case of deca-alanine [6] the corresponding quantity had the somewhat higher value of 20%.

3.3. Comparing modes from DAS and CCS analyses

Using Eq. (6) inner products between all modes from the DAS and the CCS analyses were calculated. Fig. 4 shows the result. Each symbol in the figure represents a significant correspondence between a mode from the DAS analysis and a mode from the CCS analysis. The position in the figure indicates the frequencies of these modes. Most corresponding modes occur at low frequencies. Clearly the frequencies of the DAS modes are somewhat higher than their corresponding CCS modes. The figure of 13% for the percentage difference in the total fluctuation of the protein derived from the two methods indicates that the value for the frequency of a DAS mode is, on average, greater by 13% than the frequency value of the corresponding CCS mode. The broken line in Fig. 4 is the line corresponding to this 13%

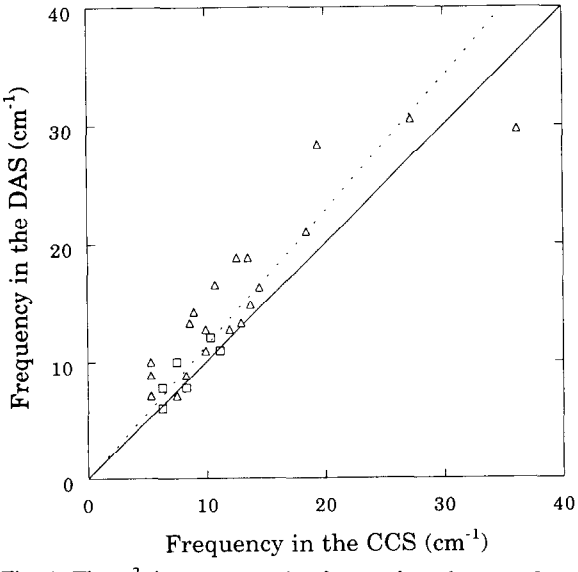

Fig. 4. The c_{ij}^2 between a pair of normal modes, one from CCS analysis, the other from the DAS. Symbols □ and △ show the c_{ij}^2 in the range 1–0.5 and 0.5–0.1, respectively. See text for explanation of broken line.

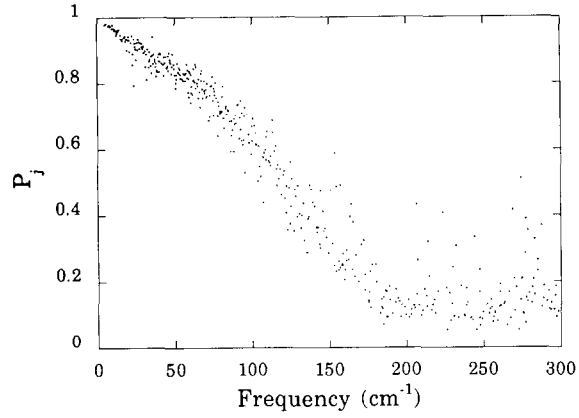

Fig. 5. Projection P_j of the jth CCS normal mode onto the DAS subspace (see Eq. (9)).

increase. As mentioned before in the case of deca-alanine the value was 20%. Fig. 5 shows the P_j values (see Eq. (9)) plotted against the mode frequency from the CCS analysis. As mentioned in Section 2 the P_j value indicates the degree to which each CCS mode can be expressed in the subspace defined by the DAS modes. Physically it tells us to what extent the motion can be represented by dihedral angle motions only. Low-frequency modes correspond to motions almost totally representable in terms of dihedral angle motions only. However, the contribution from bond angle bending and bond length stretching motions increases monotonically until they completely dominate the modes corresponding to frequencies above 200 cm^{-1}. This corresponds to the result seen in Fig. 2, where significant differences between the CCS spectra and the DAS spectrum were also seen above roughly 200 cm^{-1}. Fig. 6 shows the cumulative percentage contribution to the total mass weighted mean-square fluctuation of the protein plotted against mode frequency from the CCS analysis. It shows to what degree the low-frequency modes contribute to the total fluctuation of the protein. Modes with frequencies lower than 30 cm^{-1} contribute almost 80% of the total mass weighted mean-square fluctuation of the protein. Looking at Fig. 5 one can see that all these modes have P_j values above 0.9. Therefore one can conclude that 80% of the total mass weighted means-quare fluctuation of

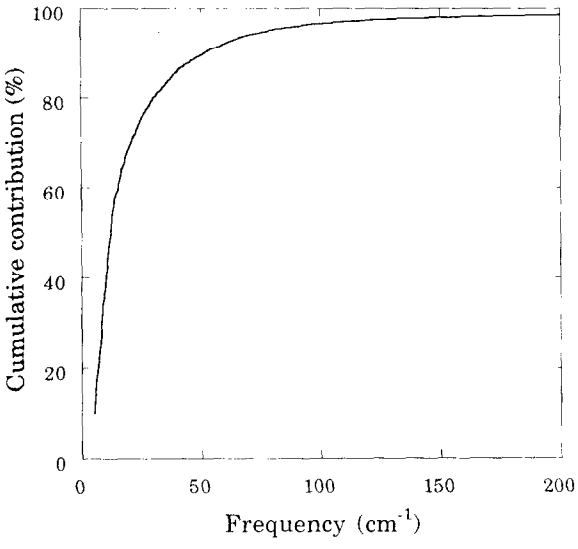

Fig. 6. Percentage of cumulative contribution of CCS modes to the total mass weighted mean-square fluctuation of the protein plotted against frequency.

the protein arises from motions involving only dihedral angles. Modes of frequencies greater than 200 cm^{-1}, which were found to be those that are dominated by bond angle bending and bond length stretching motions contribute less than 2% to this total. The intermediate case with P_j values centred on 0.5, corresponds to motions involving both dihedral angle, and bond angle and bond length fluctuations. Motions at these frequencies are not global, but span several residues.

Although it may be argued that the effect of the difference in the energy parameter sets in the two analyses is difficult to quantify, we feel sure that the results presented are, at least qualitatively, correct. Supporting this, is that, in the case of the deca-alanine analysis, where great care was taken to give common parameters identical values, qualitatively similar results were observed. The 20% difference, compared to 13% here, of the percentage increase in the RMSD value from the CCS over that found from the DAS in the case of deca-alanine, may be attributable to the different energy functions used in the CCS analyses. It is thought that the CVFF energy function used here, allows less flexibility than the AMBER [15] function used in the deca-alanine analysis.

4. Conclusions

A normal mode analysis in dihedral angle space has been compared to a normal mode analysis in Cartesian coordinate space. The CCS normal mode analysis has the advantage of being technically easier to apply than the DAS analysis. The disadvantage of the CCS normal mode analysis lies in the comparatively large number of variables that are treated when a large protein is tackled, making the energy minimization process and the diagonalization of the resulting Hessian matrix, a computationally daunting task. This is where the DAS method has an advantage. The reduction in the number of variables and consequent reduction in the computation time for the energy minimization and large reduction in the sizes of the matrices to be diagonalized, is such that even large proteins can be dealt with.

The low-frequency CCS modes (< 30 cm^{-1})can be represented almost completely by motions caused by dihedral angle motions only. Such motions account for 80% of the total mass weighted mean-square fluctuation of the protein in the CCS. Motions above 200 cm^{-1} are those dominated by bond length stretching and bond angle bending, but these contribute less than 2% to the total mean-square fluctuation. However, despite the fact that these dominant low-frequency motions can be almost totally represented by motions of only the dihedral angles, the frequencies of the DAS modes are on average 13% greater than their corresponding CCS modes. This means that the RMSF values for the low-frequency normal modes in the CCS are on average 13% greater than their DAS counterparts. It appears therefore that allowing fluctuations in bond lengths and bond angles has the indirect effect of making the dihedral angles more flexible. The direct contribution from bond angle bending and bond length stretching to the total fluctuation of the protein is in contrast very small.

Although the DAS modes have somewhat higher frequencies than their CCS counterparts, the fact that the low-frequency CCS modes are well represented by low-frequency DAS modes (see Fig. 4) means that the low-frequency modes from both analyses occupy the same subspace.

This fact is important, as it is in the subspace defined by the low-frequency modes that most of the motion occurs (see Fig. 6). In fact this is the concept of the important subspace [16]. The determination of the important subspace allows one to do normal mode X-ray refinement [16] and suggests the possibility of doing molecular dynamics simulations within this space [3]. The results here indicate that the DAS normal mode analysis can be applied to proteins too large for the CCS analysis in order to determine the important subspace.

Acknowledgments

Computations were done in computer centers at Kyoto University and at the Institute for Molecular Science. This work was supported by grants from the Ministry of Education, Science and Culture, Japan, and from the International Human Frontier Science Program. S.H. is a recipient of a European Community Science and Technology Program Fellowship for Japan.

References

[1] A. Kitao, F. Hirata and N. Go, Chem. Phys. 158 (1991) 447.
[2] S. Hayward, A. Kitao, F. Hirata and N. Go, J. Mol. Biol. 234 (1993) 1207.
[3] A. Amadei, A.B.M. Linssen and H.J.C. Berendsen, Proteins 17 (1993) 412.
[4] T. Horiuchi and N. Go, Proteins 10 (1991) 106.
[5] S. Hayward, A. Kitao and N. Go, Protein Sci. 3 (1994) 936.
[6] A. Kitao and N. Go, J. Comput. Chem. 12 (1991) 359.
[7] F.A. Momany, R.F. McGuire, A.W. Burgess and H.A. Scheraga, J. Phys. Chem. 79 (1976) 2361.
[8] Discover user guide (Biosym Technologies, 1993).
[9] M. Marquart, J. Walter, J. Deisenhofer, W. Bode and R. Huber, Acta Crystal. B 39 (1983) 480.
[10] S. Sunada and N. Go, submitted for publication.
[11] C. Eckart, Phys. Rev. 47 (1935) 552.
[12] T. Noguti and N. Go, J. Phys. Soc. Japan 52 (1983) 3283.
[13] W.L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc. 110 (1988) 1657.
[14] M.M. Teeter and D.A. Case, J. Phys. Chem. 94 (1990) 8091.
[15] S.J. Weiner, P.A. Kollman, D.T. Nguyen and D.A. Case, J. Comput. Chem. 7 (1986) 230.
[16] A. Kidera, K. Inaka, M. Matsushima and N. Go, J. Mol. Biol. 225 (1992) 477.