PROTEINS: Structure, Function, and Genetics 8:258–279 (1990)

# Normal Mode Analysis of Human Lysozyme: Study of the Relative Motion of the Two Domains and Characterization of the Harmonic Motion

**Jean-François Gibrat and Nobuhiro Gō**
*Department of Chemistry, Faculty of Science, Kyoto University, Kyoto, 606 Japan*

***ABSTRACT*** **A normal mode analysis of human lysozyme has been carried out at room temperature. Human lysozyme is an enzyme constituted of two domains separated by an active site cleft, the motion of which is thought to be relevant for biological function. This motion has been described as a hinge bending motion. McCammon et al.[1] have determined the characteristics of the hinge bending motion but they assumed a prior knowledge of the hinge axis. In this work we propose a method which is free from this assumption and determines the hinge axis and root mean square (rms) rotation angle which give the best agreement with the pattern of changes in all the distances between nonhydrogen atoms in the two domains, obtained by the normal mode analysis. The hinge axis we found is notably different from the one previously determined and goes, roughly, through the $C^{\alpha}$ 55 and $C^{\alpha}$ 76, i.e., it is located at the base of the β-sheet of the second domain. The rms value for the rotation angle is also twice as large as the previous one: 3.37°. It is shown that this hinge bending motion provides a fairly good approximation of the dynamics of human lysozyme and that the normal mode with the lowest frequency has a dominating contribution to this hinge bending motion.**

**A study of the accessible surface area of the residues within the cleft reveals that the motion does not result in a better exposure to the solvent of these residues.**

**A characterization of the thermally excited state (under the hypothesis of the harmonicity of the potential energy surface) has been done using the concept of topology of atom packing. Under this hypothesis the thermal fluctuations result only in a small change of the topology of atom packing, leading therefore to nearly elastic deformations of the protein.**

**Key words: topology of atom packing, hinge bending motion, accessible surface area, elastic deformation**



## INTRODUCTION

There is now well-documented evidence that mobility plays an important role in the biological activity of proteins.[2] A number of proteins assume two different conformations with different activities. The transition between these two states, which can be triggered by a cofactor, is thought to be due to some collective motions, that is, parts of the protein move in a relatively rigid way with respect to other parts of the molecule. An example of such a protein is the catabolite activator protein (CAP), a regulatory protein made of two domains, one binding the cofactor (cyclic AMP) and the other interacting with the DNA.[3] The two domains are mobile relative to each other. In the absence of the cofactor, the second domain does not interact with the DNA. Another example is afforded by enzymes made of two domains delineating an active site cleft. The relative motions of the two domains with the concomitant opening and closing of the cleft are also thought to be important for explaining the catalytic reaction (the best known case is maybe citrate synthase[4]).

Numerous experimental methods provide various information about the mobility of proteins: X-ray diffraction, neutron time-of-flight, spectroscopic techniques such as NMR, ESR, Mössbauer spectroscopy, and Raman spectroscopy. Spectroscopic methods give information about the distribution of low-frequency vibrations, i.e., about time or energy resolution. Until recently, X-ray diffraction (through the determination of the temperature factors: B-factors) was likely to give, in spite of experimental and theoretical difficulties, the most direct insight about the spatial resolution of protein dynamics. With the recent development of powerful NMR methods (COSY and NOESY spectroscopy), one disposes of a new tool for studying the mobility of proteins.

Computer simulations provide an alternative for

Received March 26, 1990; accepted April 16, 1990.
Address reprint requests to Professor Nobuhiro Gō, Department of Chemistry, Faculty of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto, 606 Japan.
Jean-François Gibrat's permanent address: Unité d'Ingénierie des Protéines, Bât. des Biotechnologies, INRA, 78350 Jouy-en-Josas, France.

a better understanding of the dynamics of proteins. There are mainly three methods which can be used: (1) molecular dynamics calculations,[5-8] (2) Monte Carlo simulation in the space of normal mode variables,[9] and (3) normal mode analysis.[10-14] This latter method is well suited for analyzing collective motions because the dynamics of the system is naturally expressed as a superposition of mutually independent collective variables. However, the method has the serious limitation that the potential energy, in the thermal range of fluctuations at least, must be approximated by a quadratic function of the generalized coordinates.

In this study we carried out a normal mode analysis of human lysozyme. Human lysozyme is an enzyme with antibacterial properties. It is constituted of two domains separated by an active site cleft. The X-ray structure[15] is determined with a good resolution (1.5 Å and an $R$ value of 0.187). Reliable data about the mean square fluctuations of backbone atoms deduced from the B-factors are also available,[16] which allows one to compare the experimental and theoretical results. Because the active site of the enzyme is in the cleft between the two domains, the motion of these domains is thought to be important for biological activity. It has been shown[17] that the modes with very low frequency dominate the dynamics of the system, therefore we studied the contribution of these modes, taken individually, to the motion of the protein. We paid special attention also to the relative motion of the two domains and to the resulting change in the accessible surface area of the residues pertaining to the cleft. Another feature of interest was the study of the change, if any, in the topology of atom packing when the protein undergoes motions due to the normal modes with very low frequency. This is related to the study done by Noguti and Gō[18] about the substates existing in the protein conformation space where it is shown that local deformation observed between different minima is accompanied by change in the topology of atom packing. In our case we studied the change of topology of atom packing between the minimum and thermally excited states. The topology of atom packing supplies a criterion for the characterization of the thermally excited state.

## METHOD

### General Procedure

All the atoms are represented including the aliphatic hydrogens. There are 2013 atoms in the protein. The rotatable dihedral angles are taken as independent variables, assuming standard values for the bond lengths and bond angles. This has the advantage of reducing the number of variables: 771 compared to 6033 if the cartesian coordinates are used. On the other hand with this choice it is necessary to find the set of dihedral angles such that the conformation generated best fits the X-ray coordinates (regularization step).

The theory of normal mode analysis applied to proteins has already been given in detail (see, for example, Levitt et al.[13]). Here we just give a brief outline of the method.

The first step is the location of a minimum energy conformation starting from the regularized conformation. The conformational energy is computed using ECEPP[19] energy functions (the simulation is done in vacuum). The minimization is performed using Newton's method, taking advantage of the development of an algorithm for the fast computation of the gradient and the hessian.[20-22]

The second step is the calculation of the second derivative matrix (the hessian) at the minimum point.

The third step is the calculation of the coefficient matrix for the kinetic energy when this latter is expressed as a function of the generalized coordinates.[23]

The last step is the solution of the generalized eigenvalue problem. The eigenvalues give the vibrational time scales (frequencies) and the eigenvectors give the principal axes of the quadratic. These axes are the directions in the variable space where the value of the quadratic is expressed as a sum of square without any cross terms and are therefore independent of each other.

### Expression of the Results

The results are then expressed in the following way, if $\sigma_i$ is the $i$th normal coordinate, $q_j$ the $j$th generalized coordinate (in our case one of the 771 dihedral angles) and $\mathbf{r}_a$ the position vector of the $a$th atom in the minimum energy conformation:

$$\sigma_i(t) = A_i\cos(2\pi\nu_i t + \delta_i) \tag{1}$$

where $A_i$ is the amplitude, $\delta_i$ is the phase of the $i$th normal coordinate, and $(2\pi\nu_i)^2$ is the $i$th eigenvalue previously determined.

$$\Delta q_j(t) = q_j(t) - q_j^0 = \sum_i u_{ji}\sigma_i(t) \tag{2}$$

where $u_{ji}$ is the $j$th component of the $i$th eigenvector.

$$\Delta\mathbf{r}_a(t) = \sum_j \mathbf{k}_{aj}\Delta q_j(t) = \sum_j\sum_i \mathbf{k}_{aj}u_{ij}\sigma_i(t) \tag{3}$$

where $\mathbf{k}_{aj}$ is the partial derivative of $\mathbf{r}_a$ with respect to $q_j$. An analytic expression of $\mathbf{k}_{aj}$ has been given in the ref. 23. The average value of the normal coordinate $\sigma_i$ in the thermal equilibrium is

$$\langle\sigma_i^2\rangle = \frac{k_B T}{(2\pi\nu_i)^2} \tag{4}$$

where $k_B$ is the Boltzmann constant and $T$ is the absolute temperature. Therefore the mean of the

square of the fluctuations of the cartesian coordinates of the $a$th atom is given by

$$\langle \Delta\mathbf{r}_a . \Delta\mathbf{r}_a \rangle = \sum_j \sum_k \sum_i \mathbf{k}_{aj} . \mathbf{k}_{ak} u_{ji} u_{ki} \frac{k_B T}{(2\pi\nu_i)^2} \quad (5)$$

If we consider the change of the distance between the $a$th and $b$th atoms in the protein, we have the (first-order) equation:

$$\Delta d_{ab} = (\Delta\mathbf{r}_a - \Delta\mathbf{r}_b) . \mathbf{e}_{ab} \quad (6)$$

where $e_{ab} = (\mathbf{r}_a - \mathbf{r}_b)/|\mathbf{r}_a - \mathbf{r}_b|$. The mean of the square of the change in the distance between atoms $a$ and $b$ is thus given as a sum of contributions from individual modes by

$$\langle (\Delta d_{ab})^2 \rangle = \sum_i (\Delta d_{abi})^2 \quad (7)$$

where $\Delta d_{abi}$ is a signed quantity given by

$$\Delta d_{abi} = \frac{(k_B T)^{1/2}}{(2\pi\nu_i)} \sum_j (\mathbf{k}_{aj} - \mathbf{k}_{bj}) . \mathbf{e}_{ab} u_{ji} \quad (8)$$

### Generation of a Part of the Crystal of Human Lysozyme

The purpose is to check the packing of a lysozyme molecule in the crystal. More specifically, we are interested by the environment of loop 40–60 (first loop of the second domain) which shows during the minimization an unusual deviation from the X-ray structure.

Human lysozyme crystallizes in the orthorhombic space group $P2_12_12_1$ with cell dimensions $a = 57.1$ Å, $b = 61.0$ Å, and $c = 32.9$ Å. There are four molecules in the unit cell, whose coordinates are given by M1 $(x,y,z)$, M2 $(a/2-x,b-y,c/2+z$, M3 $(a-x,b/2+y,c/2-z)$, and M4 $(a/2+x,b/2-y,c-z)$. We built a part of the crystal, i.e., the original cell (cell 0), cell 1 (translation $+c$), cell 2 (translations $+c$ and $-a$), cell 3 (translations $+c$, $-a$, and $-b$), cell 4 (translations $+c$ and $-b$), cell 5 (translation $-a$), cell 6 (translations $-a$ and $-b$), cell 7 (translation $-b$), cell 8 (translation $-c$), cell 9 (translations $-c$ and $-a$), cell 10 (translations $-c$, $-a$, $-b$), and cell 11 (translations $-c$ and $-b$). Only these 11 cells (among the 26 surrounding the initial cell) are considered, because we are interested in the molecules which are in the immediate vicinity of molecule M1 in the original cell. Then, we check the distances between atoms of the loop consisting of residues 40–60 of M1 in the original cell (in the following we will use the compact notation: R40-60M1C0) and the atoms of other molecules in these 11 cells. We consider only atoms which are within a sphere of 6 Å radius centered on each atom of the loop.

### Map Showing the Change of the Distance Between All the Possible Pairs of Residues Due to a Single Normal Mode

The distance between two residues is defined as the average of the distances between corresponding atoms of the main chain (N, $C^\alpha$, C′, O). The change in the distance is calculated using the signed quantity defined by Eq. (8). The change in the distance between two residues is then calculated as the average change between corresponding atoms of the main chain. This quantity is normalized by dividing by the average change of all the possible distances in the molecule. If the magnitude of this new quantity is greater than 1 the position in the map corresponding to the distance in question is marked with a '+' or a '−' according to the sign of $\Delta d_{abi}$. It has to be kept in mind that the atoms are oscillating around their equilibrium position. For the sake of the drawing one direction has been arbitrarily chosen, in which some atoms are going further apart and some are getting closer. Obviously, half a period later the opposite motion will be observed, i.e., the "+" will be changed to a "−" and conversely.

### Motion of the Two Domains

In human lysozyme, as in several other enzymes, the active site is located in a cleft between the two domains. Crystallographic studies, in a number of case, have shown the existence of two different conformations: "open" when the enzyme is free and "closed" when a substrate is bound. Therefore the motion of these domains is thought to be important for the biological activity. The motion of the two domains has been described as a hinge bending motion, i.e., the two domains move as rigid bodies around an axis of rotation.

Theoretical studies have been carried out for the purpose of analyzing the hinge bending motion. In the work of McCammon et al.[1] the molecule (hen egg white lysozyme) is divided into two domains and the position of the hinge axis is guessed by examining the crystal structure. Then an adiabatic potential is determined by holding the rotation angle at a series of values and allowing the positions of atoms in the displaced domains to adjust by minimizing the energy. This allows the determination of the force constant associated with the relative motion of the lobes. This force constant is then used in the Langevin equation for a damped harmonic oscillator. In this model, the first domain is constituted by residues 5–36 and 98–129 and the second by residues 40–94 and the hinge axis is chosen as passing through $C^\alpha$ 38 and $C^\alpha$ 97. The rms value of the rotation angle they found is 1.6°. In a subsequent study, Brooks and Karplus[24] used a method which extracts a normal mode or modes of the full second-derivative matrix overlapping most with a starting model for the motion. The starting model consists in the rigid-body hinge bending motion previously described. They found two modes, one with the frequency 3.56 $cm^{-1}$ has an overlap of 0.87 and the other with a frequency of 4.78 $cm^{-1}$ has an overlap of 0.5 with the rigid-body motion. This method provides a description of the motion which is indepen-

dent of the details of the initial guess. Although this technique does not allow a precise determination of the hinge axis, they estimate that the hinge axis goes roughly through the base of the β-sheet at Thr-40 and the disulfide bridge between Cys-76 and Cys-94. In this model the first domain would therefore correspond to the residues 1–39 and 89–129 and the second domain to the residues 40–88.

Few experimental data are available on the dynamics of the hinge bending. However, Sternberg et al.[25] showed that the particular hinge axis found by McCammon et al.[1] does not fit well the data obtained by X-ray diffraction (B-factors).

In order to test the hypothesis of a hinge axis, we propose here a method which determines the hinge bending axis and mean angle which fit best the data obtained by the normal mode analysis. This method does not presume any prior knowledge of the position of the hinge axis. The only assumption is the division of the protein into two, well-characterized, domains.

A hinge bending motion can be described by six parameters: an angle of rotation around the hinge axis, the three coordinates of a point on the axis, and, for example, two angles expressing the direction of the hinge axis. In such a motion, all the possible distances between atoms in one domain and atoms in the other domain change by a given amount. The normal mode analysis allows us to determine the mean of the square of the change in the distance between two atoms due to one mode or due to the combination of several modes [Eq. (7) above]. When all the distances between the two domains are considered this leads to a specific pattern of change in the distances. The basic idea of the method is, therefore, to determine the set of parameters describing the hinge bending motion which gives a pattern of change in all the distances between the two domains which fits best the pattern of change obtained by the normal mode analysis.

For this purpose we use a least-squares minimization method, the Levenberg–Marguardt–Morrison method, implemented in the SSL II mathematical library (subroutine DNOLG1). The problem consists in finding the minimum of the following target function:

$$I(\mathbf{x}) = \frac{1}{2}\sum_a\sum_b \{[\Delta d_{ab}(\mathbf{x})]^2 - \langle(\Delta d_{ab})^2\rangle\}^2 \qquad (9)$$

where $\mathbf{x}$ stands for the six parameters necessary to describe the hinge bending motion. $[\Delta d_{ab}(\mathbf{x})]^2$ is the square of the change in the distance between a pair of atoms, $a$ (first domain) and $b$ (second domain), calculated for the hinge bending motion defined by a given set of parameters $\mathbf{x}$. $\langle(\Delta d_{ab})^2\rangle$ is the mean of the square of the variation in the distance of the same pair of atoms, $a$ and $b$, obtained by the normal analysis [Eq. (7)].

In other words, we consider the change in the distances obtained with the normal mode analysis as "experimental points." Using a model (here the hinge bending motion) which is described by a number of parameters, we can calculate similar points. Then we have to find the set of these parameters which minimizes the sum of the square of the difference between the calculated and "experimental" points. The first domain consists of residues 1–40 and 88–130 and the second domain consists of residues 41–87. We consider only the distances between the heavy atoms in both domains and, thus, there are about 250,000 such points.

As Eq. (7) indicates, the variation in the distance of a pair of atoms is given as a sum of contribution from individual normal modes. In the following, we will examine five different cases. In the first four cases, the contribution from just one normal mode is considered (the first four modes). In these cases we ask the following questions by minimizing the target function of Eq. (9). If such a mode can be approximated by a hinge bending motion between the two domains, where is the hinge axis, and how good is the approximation? In the fifth case, contributions from the first 10 lowest frequency modes are considered. This provides a reasonable approximation of contributions from all the modes. Then the questions we are asking are: What is the effect of the combination of different modes? Can the harmonic motion of the protein molecule be approximated by a hinge bending motion? If so, what are the characteristics of the hinge bending motion and how are they related to the characteristics of the hinge bending motion found for individual normal modes? Finally, as for the first four cases, we have to estimate how well the model fits the data.

Before answering these questions, we have to assess the ability of the least-square minimization described here to find the hinge axis. For that purpose we created some test data corresponding to the change in the distances between the two domains due to a rotation of 3° around the hinge axis going through the $C^\alpha$s 41 and 87. Starting with different points in the space of parameters, the method was able to find exactly the hinge axis and the rotation angle, except one, when the starting point was such that the axis was perpendicular to the true axis and going through a point located in the middle of the line joining $C^\alpha$ 48 and $C^\alpha$ 110 (that is in the middle of the cleft at the opposite end with respect to the hinge axis). The value of the function at the minimum point was virtually zero, except for the latter case.

Our model is based on the assumption that we know precisely the boundaries of the two domains, a fact which might be arguable. Therefore, we probed the effect of a wrong choice for the limits of the two domains. The molecule was divided into a first domain constituted of the residues 3–37 and 98–130

and a second domain constituted of the residues 1–2 and 38–97. With this partition of the molecule, data similar to the one above were calculated for the hinge axis going through $C^{\alpha}$s 38 and 97 and for the rotation angle of 2°. This is similar to the hinge axis used by McCammon et al.[1] and corresponds roughly to their partition of the molecule. A minimum was then searched for, but using the initial partition of the molecule (domain 1: residues 1–40 and 88–130 and domain 2: residues 41–87). Again starting from different points the method found a single minimum. At this point the value of the function was 36 Å$^4$. The rotation angle was 1.81° and the direction of the hinge axis found made an angle of 7° with the real axis.

In the light of these results, it appears that the method is able to find, without great difficulties, the hinge axis (except if a very unreasonable starting point is chosen). Besides, with the exception previously mentioned which corresponds to a thin cone of three degrees around the direction perpendicular to the real hinge axis, all the other starting points lead to a single minimum.

### Change in the Topology of Atom Packing

The change in the topology of atom packing has been defined by Noguti and Gō[18] in their study of the minima populating the conformational energy surface. This change is defined between two conformations corresponding to different minima. All the pairs of atoms in contact (PACs) in the two structures are determined. A pair of atoms is considered as being in contact if the distance between the two atoms is less than 1.2 time the sum of their van der Walls radii if they are not engaged in a hydrogen bond. If such, the two atoms are considered in contact if their distance is less than 1.2 time the distance corresponding to the minimum for the hydrogen bond potential function (12–10 potential). The union of the two set of PACs is then taken. The subset of PACs for which the distance between the two atoms changes by more than 0.5 Å in the two structures defines a change in the topology of atom packing of the molecule. In our case we consider the change of topology of atom packing between the minimum and the structures resulting from fluctuations in the atomic positions due to the 10 lowest normal modes taken separately on the one hand and to the combined action of these modes on the other hand. In order to estimate the change of topology of atom packing due to the thermal motions of the normal modes, we computed, for all the pairs of atoms in contact in the minimized structure or liable to be brought in contact by the thermal fluctuations, the root mean square of the change in the corresponding distances using Eq. (7). When the effect of an individual mode is studied, only the contribution of the corresponding mode is considered in Eq. (7), i.e., no summation is carried out. If $\Delta d_{ab}$ is larger than 0.5 Å, this pair of atoms is regarded as incurring a change in the topology of atom packing.

## RESULTS

### Regularization and Minimization

Here, we just give a brief description of the principal features of the three-dimensional structure of human lysozyme. Human lysozyme is made of two domains delineating an active site cleft. The first domain (residues 1–40 and 88–130) is a very solid one made of four α-helices, [helix A (5–14), helix B (25–35), helix C (90–99), and helix D (110–116)] and a short parallel β-sheet [strand 1 (1–3) and strand 2 (38–40)]. This domain is cross-linked by two disulfide bridges Cys-6–Cys-128 and Cys-30–Cys-116. The second domain (residues 41–87) is principally constituted by two loops which form a three-stranded antiparallel β-sheet [strand 1 (42–46), strand 2 (51–55) and strand 3 (59–61)]. The second loop is somewhat rigidified by an internal disulfide bridge (Cys-65–Cys-81). The last disulfide bridge (Cys-77–Cys-99) joins the two domains. The active site cleft is mainly constituted by three clusters of residues in the second domain: 35, 36, 37 then 44, 46, 51, 53, 57, 58, 59, 60, 63, 64, and 74, 75 and an almost continuous region in the first domain: 99, 100, 101, 102, 103, 104, 108, 109, 110 (114), 115.

After regularization and minimization, the root mean square displacement (rmsd) of the $C^{\alpha}$ atoms in the generated structure compared with the X-ray structure is 1.08 Å. If all the heavy atoms are considered, the rmsd is 1.54 Å. However, a closer examination of the structure reveals that the variation from the X-ray structure is not homogeneous in the whole structure.

If the second domain is excluded from the best fit calculation, the rmsd for the $C^{\alpha}$ is 0.56 Å (in that case, if all the $C^{\alpha}$s are included in the rmsd calculation, the value is 1.35 Å). If the first domain is excluded from the best fit calculation, the rmsd for the $C^{\alpha}$ is 1.02 Å (1.94 Å if all the $C^{\alpha}$ are included). Therefore the first domain does not show a large deviation from the X-ray structure. On the other hand the second domain exhibits quite a large deviation. Figure 1 displays the distances between corresponding $C^{\alpha}$s in the two structures after that the rotation which gives the best fit on the first domain has been applied to the whole structure. The end of the first loop in the second domain shows a large deviation, up to 6 Å. The second loop shows also some deviation but to a lesser extent (about 2 Å for residues 66 and 81). There is only a slight motion in the first domain around residue 120. Figure 2 shows a stereo view of the two structures illustrating these features.

Because the second domain is made of a short β-sheet containing three strands it is interesting to pay attention to the behavior of the hydrogen bond pattern during the minimization (see Table I). As a

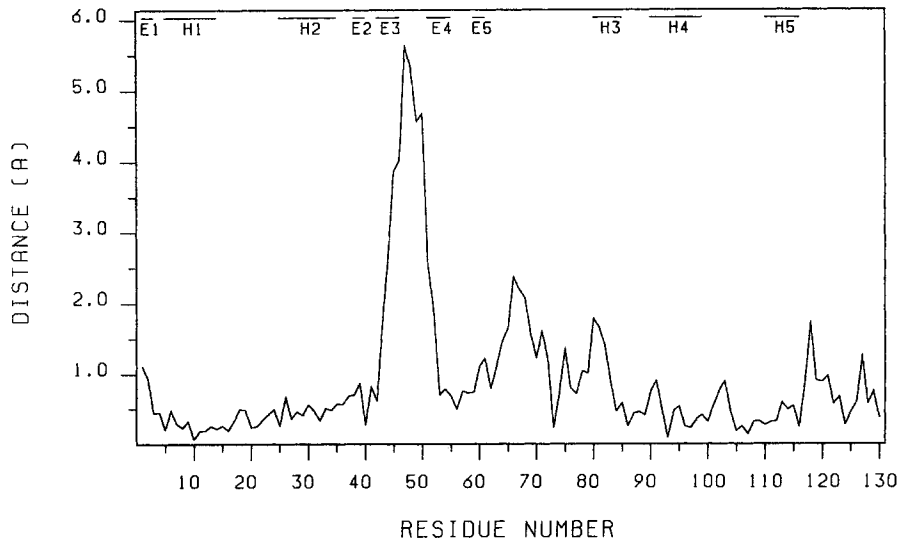


Fig. 1. Plot of distances (Å) between the corresponding $C^{\alpha}$ in the X-ray structure and the minimized one. The minimized structure has been rotated in such a way that the first domain (residues 1–40 and 88–130) has the best fit (0.56 Å) with the first domain of the X-ray structure. Elements of secondary structure are shown on the top of the drawing (E: β-strand and H: α-helix).

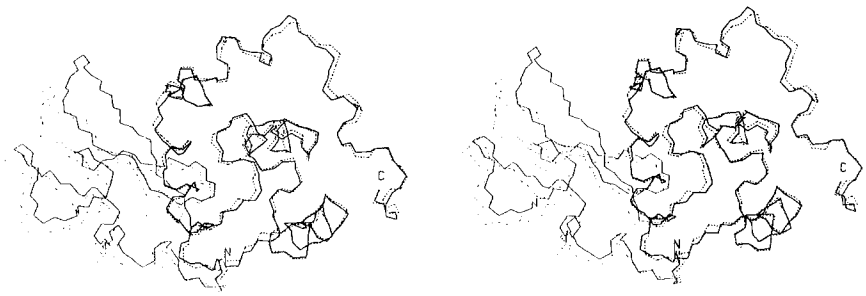

Fig. 2. Stereo drawing of the X-ray structure (dotted line) and the minimized one (solid line) after the rotation which gives the best fit between the first domains has been applied to the latter. For both conformations, the first domain is drawn with a thicker stroke.

general rule, the distance usually increases in the minimized structure. Especially at the beginning of the sheet between the first two strands (residues 42–55 and 44–53) and at the end of it between the last two strands (residues 52–61). The hydrogen bond pattern is somehow weakened but by no mean destroyed even for the large scale motion during the minimization process.

Such a large motion of the first loop in the second domain with respect to the X-ray structure is quite unusual and needs some explanation. Among all the possible causes to this phenomenon, the effect of crystal packing appears to be the most natural one. This prompted us to check the packing of the molecules in the crystal.

We built a part of the crystal and calculated the distances between atoms of the loop of the molecule M1 in the original cell (simply referred to the loop hereinafter) with all the other molecules in the same cell and in the surrounding cells (see method section). The minimum distances with each cell which have been found are given in Table II. In the X-ray structure there are 257 pairs of atoms with a distance less than 6 Å with M2 of cell 4, 3 pairs with M4 of cell 6, and 2 pairs with M2 of cell 7. Most of the pairs with M2 of cell 4 involve the main chain of the loop with the side chains of Arg-115 and Arg-119 and the side chain of Arg-50 in the loop with the main chain and side chain of residues located between Arg-113 and Arg-119. There are 7 pairs for which the distance is less than $r_{min}$ (the distance corresponding to the minimum for the interaction between 2 atoms using ECEPP parameters), the smallest distance being 2.81 Å between O of Ala-47 in the loop and NH1 of Arg-119 of M2C4 (the corresponding $r_{min}$ is 3.32 Å).

The same kind of calculation has been done replacing the X-ray coordinates by the ones obtained after the minimization. The minimized structure has been rotated in such a way that the first domain

**TABLE I. Change in the Hydrogen Bond Pattern During the Minimization**

| Hydrogen bond pattern | Minimum conf. | X-ray conf.* |
|---|---|---|
| 52 CO—NH 61 | 3.30 | 2.90 |
| 54 NH—CO 59 | 3.02 | 2.90 |
| 54 CO—NH 59 | 2.83 | 3.04 |
| 42 CO—NH 55 | 3.43 | 2.71 |
| 44 NH—CO 53 | 3.46 | 2.84 |
| 44 CO—NH 53 | 2.82 | 2.75 |
| 46 NH—CO 51 | 2.90 | 2.79 |
| 46 CO—NH 50 | 2.84 | 2.74† |

*Distance in Å between the oxygen and the nitrogen in the β-sheet of the second domain.
†Hydrogen bond in the β-turn at the end of the first loop.

**TABLE II. Minimum Distances (Å) Between Atoms of Loop 40–60 in M1 of the Original Cell and Atoms in the Other Molecules in the Same Cell and in the Surrounding Cells**

| | $D_{min}$ | |
|---|---|---|
| Cell | X-ray structure | Min. structure |
| 0 | >10.0* | >10.0 |
| 1 | >10.0 | 9.2 |
| 2 | 7.0 | >10.0 |
| 3 | >10.0 | >10.0 |
| 4 | 2.8 | 1.9 |
| 5 | 8.6 | 8.4 |
| 6 | 4.3 | 5.2 |
| 7 | 5.1 | 4.9 |
| 8 | >10.0 | 9.9 |
| 9 | >10.0 | >10.0 |
| 10 | >10.0 | >10.0 |
| 11 | >10.0 | >10.0 |

*The minimum distance is larger than 10.0 Å.

has the best fit (0.56 Å as previously mentioned) with the X-ray coordinates. For the sake of comparison the hydrogen atoms were not included in the calculation.

The minimum distances are shown in Table II. There are 164 pairs of atoms with a distance less than 6 Å with M2 of cell 4, 7 with M4 of cell 6, and 13 with M2 of cell 7. For the minimized structure, most of the pairs involve now the residue Arg-50 in the loop with the side chains of Arg-115 and Arg-119 (as well as the side chain of Asn-114 and Trp-34 to a lesser extend) of M2 in cell 4. There are 20 pairs for which the distance is less than $r_{min}$ including a serious steric interaction between NH1 of Arg-115 and O of Gly-48 ($d = 1.87$ Å, $r_{min} = 3.32$ Å).

Figure 3a shows a stereo view of the area around the loop in the crystal, i.e., the residues 40–60 of the loop in the original cell, the residues 110–122 of M2 in cell 4, the residues 84–90 of M4 in cell 6, and the residues 100–106 of M2 in cell 7. The minimized structure is shown with a dotted line. According to this figure the main chain of the loop in the X-ray structure appears to be trapped between the two side chains of Arg-115 and Arg-119 of M2 in cell 4, the side chain of Arg-50 belonging to the loop lies outside. In the minimized structure the main chain of the loop is no longer between the two arginines, but instead it is replaced by the side chain of Arg-50 (see the simplified drawing in Fig. 3b). This motion might involve some serious steric constrains as well as some strong electrostatic interactions in the crystal. It is likely, therefore, that the change in the conformation of the loop during the minimization is genuine and corresponds to the conformation when the protein is not densely packed in the crystal. In this respect it is interesting to notice that among the residues of the first domain (see Fig. 1) residue 119 has the largest displacement.

### Frequency Histogram

Figure 4 shows the frequency distribution, the number of modes in each interval of 5 $cm^{-1}$. These frequencies appear to be grouped in 4 clusters:0–200, 200–450, 450–675, and 850–950 $cm^{-1}$. The lowest frequency is 3.72 $cm^{-1}$ and the highest is 954.90 $cm^{-1}$.

As for the three last clusters, two points have to be noticed. First, the approximation which consists in fixing the bond angles and the bond lengths becomes poor when the frequency of the normal modes increases. Second, for frequencies larger than 400 $cm^{-1}$ the agreement between the classical dynamics and the quantum mechanical dynamics is decreasing. For 1000 $cm^{-1}$, for example, the mean square fluctuations for the normal variables computed with the classical dynamics approximation is about 2.5 time smaller than it ought to be if determined by the quantum mechanical dynamics.

### Fluctuations of Dihedral Angles and Cartesian Coordinates

Table III presents the average root mean square fluctuations for the dihedral angles and position of the $C^{\alpha}$s.

Figure 5 shows the rms fluctuations for the $\chi$ angles. The average rms fluctuations for the $\chi$ angles are only slightly higher than the average rms fluctuation values for the $\phi$ and $\psi$ angles. However, some $\chi$ angles show large rms fluctuations. The most striking case is $\chi_2$ of Asp-87 which shows rms fluctuations of 45°. An analysis of the peaks with a rms fluctuation greater than 20° shows that these peaks correspond to the following torsional angles: $\chi 3$ of Glu and $\chi_2$ of Asp (i.e., rotation of the carboxyl group), $\chi_2$ of Ser and Thr and $\chi_3$ of Tyr (i.e., rotation of the hydrogen atom of the hydroxyl group), $\chi_2$ of Asn-114 (rotation of the amide group). Surprisingly the dihedral angles belonging to larger side chains (Arg, Lys, etc.), i.e., $\chi_4$ to $\chi_7$, do not show large fluctuations. The large fluctuations occur rather for the dihedral angles $\chi 2$. Another surprising result is the

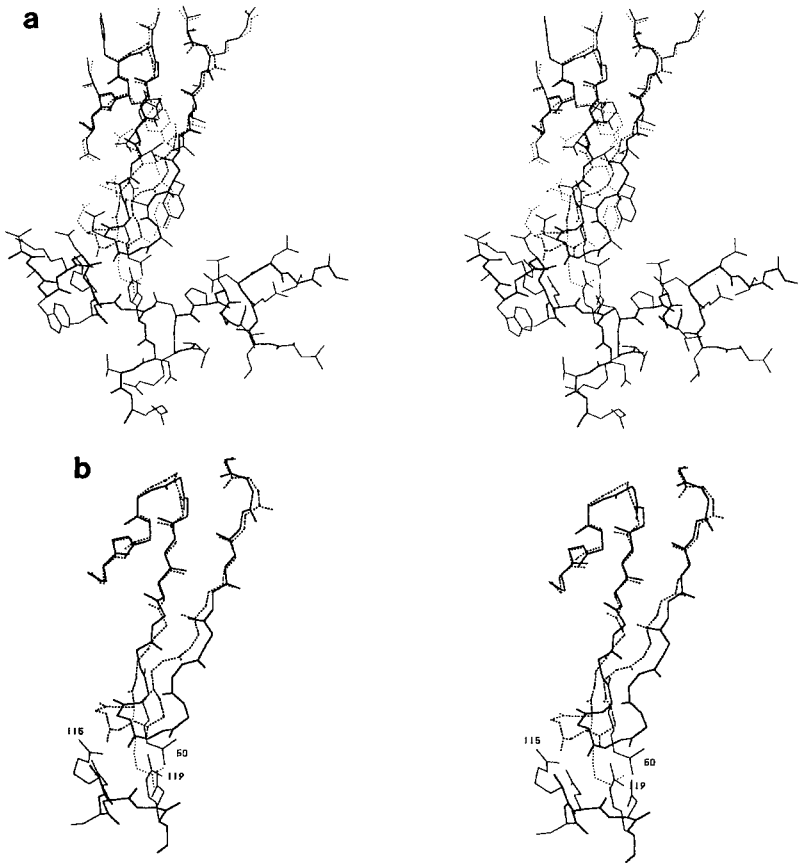


Fig. 3. (a) Stereo drawings of loop 40–60 in the X-ray structure (solid line) and in the minimized structure (dotted line) as well as fragments of other molecules in the crystal which are close (see text). (b) A simplified view of the previous drawing, only the side chain of Arg-50 and the main chain of the loop and the segment between Arg-115 and Arg-119 of molecule 2 in cell 4 are shown.

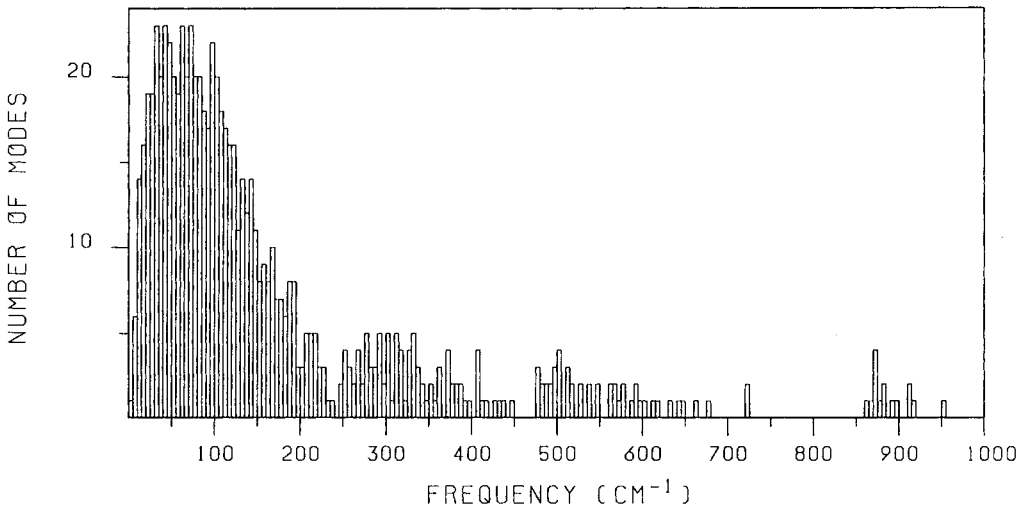


Fig. 4. Frequency distribution. Number of modes in each interval of 5 $cm^{-1}$.

fact that, except for the $\chi_2$s and to a lesser extent the $\chi_3$s, the average rms fluctuations for a given type of angle $\chi$ are approximatively constant.

It has been shown[17] that the mass-weighted mean-square displacement of the atoms is given by a sum of contribution from each normal mode, each

**TABLE III. General Results of the Normal Mode Analysis: Average Root Mean Square Fluctuations**

| | |
|---|---|
| $C^\alpha$ atoms | 0.4Å |
| $\phi$ angles | 8.3° |
| $\psi$ angles | 8.4° |
| $\omega$ angles | 5.1° |
| $\chi$ angles | 9.7° |

contribution being proportional to the inverse of the square of its frequency.

Figure 6a shows a related plot for the dihedral angles. This figure presents the average rms fluctuations for all the torsional bond angles as a function of the frequency of the normal modes. Up to the 582nd normal mode the average rms fluctuations for the angles slowly decreases. After that the behavior of the curve is more erratic with very sharp peaks such as the ones corresponding to normal modes 582, 642, 643, 645, 673 and an overall increase in the average level. This zone corresponds to normal modes with a frequency greater than 200 $cm^{-1}$ and one has to bear in mind that the approximation which consists of fixing the bond lengths and bond angles becomes poor for such frequencies. However, in order to explain this behavior we plotted the same curve but only for the backbone angles (see Fig. 6b). Up to the 575th normal mode the curve has roughly the same shape. After that point, the contribution of the normal modes to the fluctuations of the backbone angles is virtually nonexistent with the exception of the two peaks due to normal modes 669 and 723. The former peak corresponds to the fluctuations of the first angle in the molecule ($\phi$ of Lys-1) and the latter peak corresponds to the fluctuations of the last angle in the molecule ("$\omega$" of Val-130). A study of the normal modes after the 600th one shows that they are mainly responsible for the fluctuations of a few $\chi$ angles, i.e., only a few $\chi$ angles amount for more than 90% of the mean square (ms) fluctuations due to these modes.

This result prompted us to evaluate the number of heavy atoms which similarly contribute 95% of the ms fluctuations due to a given normal mode (see Fig. 7). All the atoms have been previously ordered according to their contributions to the total ms fluctuations.

With the exception of the first mode which involves only 200 heavy atoms, the first 400 normal modes involved on average about 325 atoms, i.e., one-third of the heavy atoms in the protein. After that, up to the 600th normal mode, this amount decreases regularly and after the 600th mode only a few atoms are involved. In the molecule, the distribution of heavy atoms is the following: 13% of $C^\alpha$ atoms, 38% of backbone atoms (including the $C^\alpha$s), and 62% of side chain atoms. On average, up to the 400th normal mode, the heavy atoms contributing to 95% of the ms fluctuations are distributed in the following way: 9.9 ± 1.0% of $C^\alpha$ atoms, 33.1 ± 1.7% of backbone atoms, and 66.9 ± 1.7% of side chain atoms. That is, the side chain atoms have a slight tendency to be more mobile than the backbone atoms. However, as shown by the rather small standard deviation, each of these modes involves about the same percentage of side chain and main chain atoms.

In order to obtain a more detailed picture, we plotted the contribution of the 10 normal modes having the lowest frequencies to the rms fluctuations of the $C_\alpha$ (Fig. 8a–k).

1. Normal mode #1: [3.72 $cm^{-1}$] (Fig. 8a) results in a distinct pattern of fluctuations. This mode accounts for about two-thirds of the main peak ($C^\alpha$45–$C^\alpha$51) and to a lesser extent to the rms fluctuations in the area $C^\alpha$110–$C^\alpha$120.
2. Normal mode #2: [5.29 $cm^{-1}$] (Fig. 8b). This mode contributes also to the main peak (about one-third) and to the peak in the area $C^\alpha$65–$C^\alpha$75.
3. Normal mode #3: [6.17 $cm^{-1}$] (Fig. 8c). This normal mode also results in distinct fluctuations in the area $C^\alpha$100–$C^\alpha$120.
4. Normal mode #4: [6.41 $cm^{-1}$] (Fig. 8d). This normal mode contributes slightly to the area $C^\alpha$45–$C^\alpha$51, $C^\alpha$70–80, and $C^\alpha$100–$C^\alpha$105.
5. Normal mode #5: [7.38 $cm^{-1}$] (Fig. 8e). The main contribution of this mode occurs in the region $C^\alpha$100–$C^\alpha$105.
6. The other normal modes (Fig. 8f–j) do not show particular features. Their contributions are relatively modest.

Figure 8k shows the rms fluctuations contributed by the 10 lowest normal modes previously mentioned as well as the rms fluctuations due to all the normal modes. Table IV indicates what percentage the sum of the ms fluctuations due to these modes represents with respect to the overall ms fluctuations. As pointed out by Levitt et al.[13] the sum of the fluctuations contributed by only a few lowest normal modes explains well the shape of the curve due to all the normal modes. However, no one of the individual normal modes has all the features of the complete curve. In fact the shape of the curve is given by the lowest normal modes, each contributing certain patterns (or parts of patterns). Normal modes with higher frequency contribute to background fluctuations.

It is interesting to see which angles are responsible for the $C^\alpha$ fluctuations, especially the fluctuations with distinct patterns due to the 5 lowest normal modes. Figures 9 show the rms fluctuations of $\phi$ and $\psi$ in the same individual normal modes.

1. Figure 9a: The fluctuations of $\phi$ and $\psi$ are mainly limited to area 45–53 with a large peak for $\psi$45. Fluctuating angles are therefore located in the

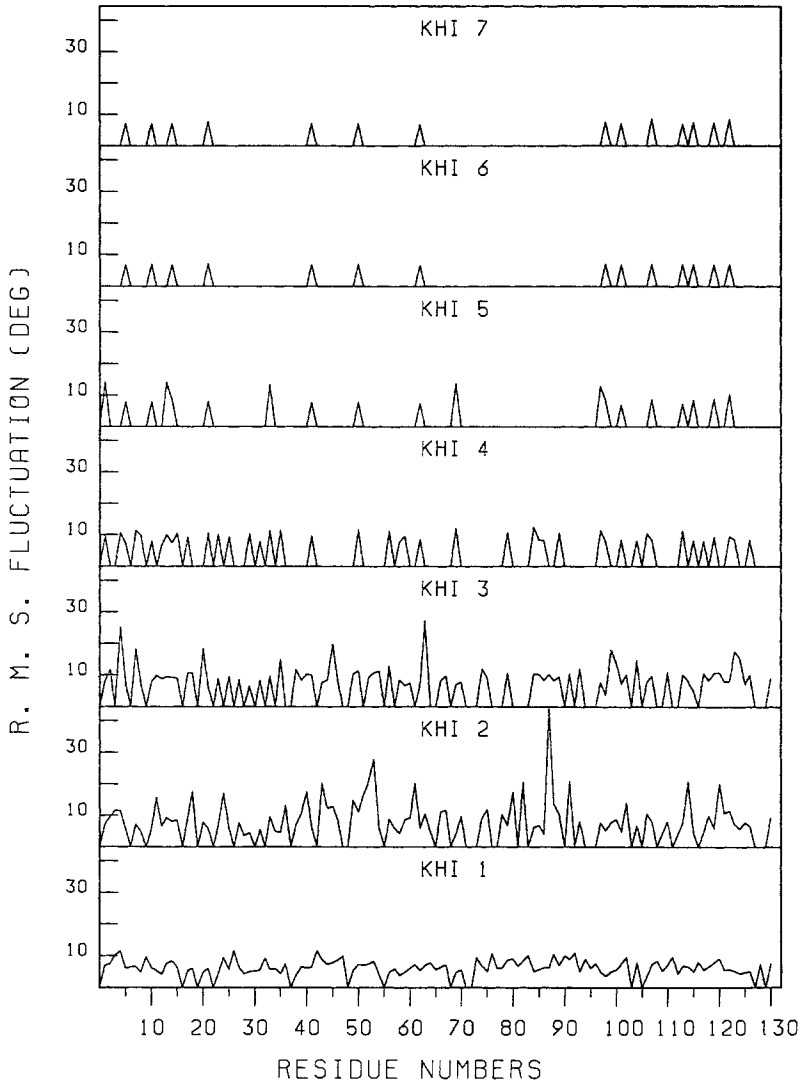


Fig. 5. Rms fluctuations (deg) of the $\chi$ angles.

same zone as fluctuating $C^{\alpha}$s. The fluctuating zone for angles appears only slightly broader. This localization of angle fluctuations implies some kind of compensation.

2. Figure 9b: The fluctuations of $\phi$ and $\psi$ angles are not so well localized in this mode. There is still several peaks in region 45–55, with again a large peak for $\psi$45, which are likely responsible for the peak of $C^{\alpha}$ fluctuations observed at 46–50. Other peaks, less important, are observed at $\psi$59, $\phi$60, and $\phi$68, which may contribute to the second smaller peak in the $C^{\alpha}$ fluctuations plot for this mode.

3. Figure 9c: For this mode the fluctuations are extremely localized at $\psi$109 and $\phi$110, which correspond exactly to the highest peak in $C^{\alpha}$ fluctuations. However in the $C^{\alpha}$ rms fluctuations plot there is a broad region fluctuating (from $C^{\alpha}$103 to $C^{\alpha}$120) with several peaks and only a sharp peak in the $\phi$ and $\psi$ plot.

4. Figure 9d: The fluctuations are quite small except a thin peak at $\psi$2 and $\phi$3. There is a group of peaks (a massif) in region 45–65 which corresponds on the $C^{\alpha}$ fluctuations plot to a broad peak.

5. Figure 9e: Here again the fluctuations are small. A broad band between 90 and 110 corresponds only to a relatively thin peak on the $C^{\alpha}$ plot around $C^{\alpha}$104.

6. Figures 9f–i: No peculiar feature to notice.

7. Figure 9j: In this plot one can notice the presence of a thin peak at $\psi$68 and $\phi$68 with a great amplitude. The plot of $\omega$ fluctuations also shows a peak at $C^{\alpha}$68 (data not shown). Curiously enough the fluctuations plot does not show any peculiar feature (there is a very small peak for $C^{\alpha}$69).

The fluctuations of $\phi$ and $\psi$ angles due to the lowest normal modes appear also to show some patterns. However, the phenomenon is less pronounced

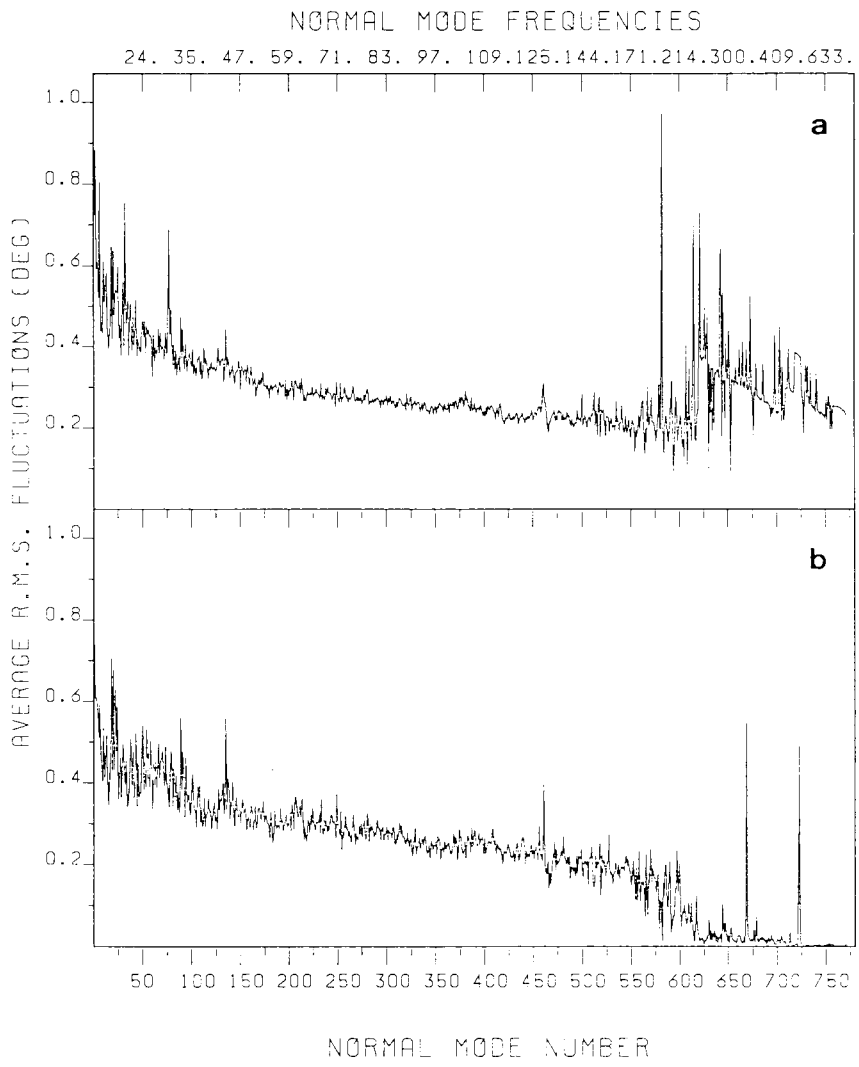


Fig. 6. (**a**) Average mean square fluctuations of all the dihedral angles in the molecule (deg$^2$) due to each normal mode. (**b**) Similar to **a** but only the backbone angles are considered.

in this case than for the $C^\alpha$ fluctuations. Besides the values of the fluctuations are, relatively, smaller than the corresponding $C^\alpha$ fluctuations due to the lowest modes. For example, the largest value for the angle fluctuations due to the 10 lowest modes is about 7° which is less than the average rms fluctuation for $\phi$ or $\psi$ angles. Each normal mode does not contribute a well-characterized pattern to the general curve (data not shown).

It has been noticed[13,14] that secondary structure elements correspond usually to zones of low fluctuations. Here this is clearly true for the five α-helices which have rms fluctuations below the average rms fluctuations (see Fig. 8k). This is not true for the β-strands with the exception of strands E1 and E2 (residues 1–3 and 38–40) which form the β-sheet in the first domain. The three other strands in the second domain show large rms fluctuations.

## Comparison With Experimental Study

Artymiuk et al.[16] carried out a crystallographic study of the dynamic properties of lysozymes. They took advantage of the determination of the structures of hen egg white lysozyme (HEWL)[26] and human lysozyme (HL)[15] to a good resolution for doing a thorough analysis of the B-factors (temperature factors). These B-factors are related to the mean square amplitude of harmonic displacement. However, as they pointed out, there are several experimental difficulties in obtaining accurate B-factors. Theoretical problems also arise in relating the observed B-factors to true atomic motions. The fact that they disposed of two structures which crystallize in different space groups, for which the data have been collected and corrected using different procedures at different resolutions, allowed them to

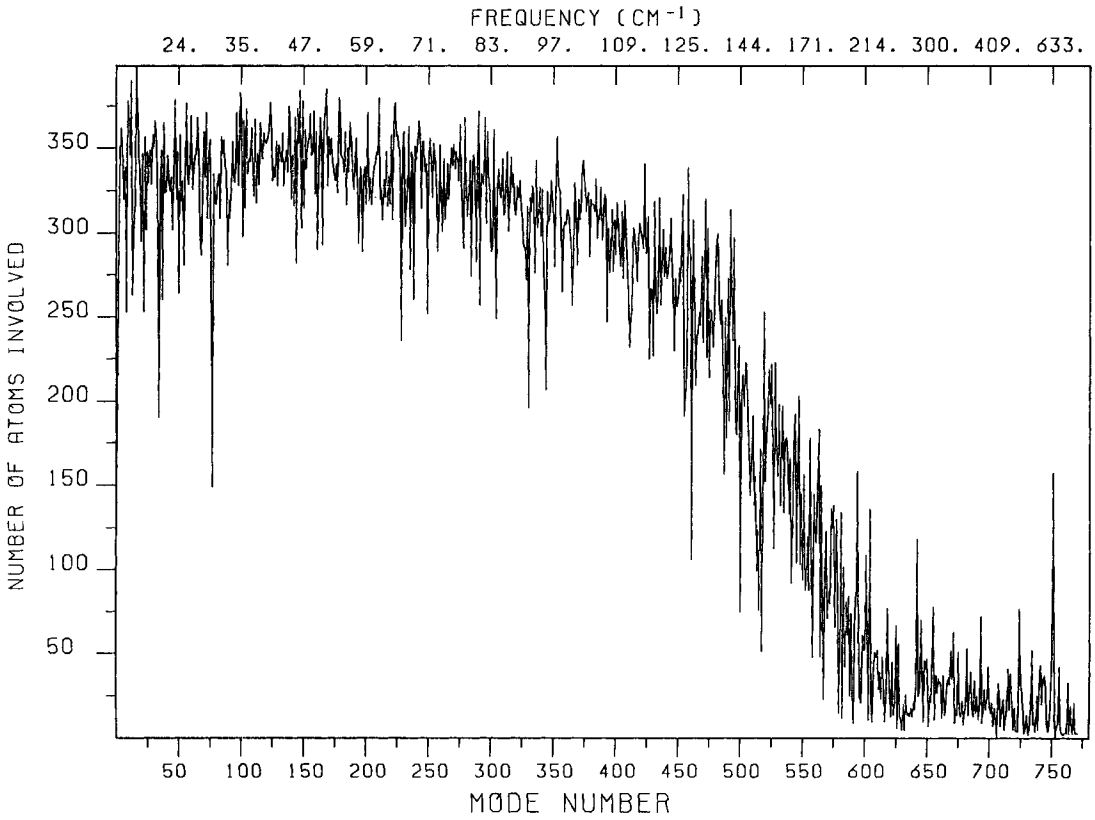


Fig. 7. Number of heavy atoms necessary for amounting to 95% of the mean square fluctuations due to a given normal mode.

postulate that the effects of experimental errors and crystal packing are not so drastic and that it is possible to relate B-factors to characteristic molecular properties.

Figure 10 shows the average mean square fluctuations for the N, $C^{\alpha}$, $C^{\beta}$, C′, and O atoms for each residue at 300 K calculated by the normal mode analysis and deduced from the B-factors.

A comparison of the two curves reveals a general agreement in the position of the peaks with a small discrepancy in region 118–125. The ms fluctuations in our simulation are usually smaller than the ms fluctuations deduced from the B-factors with the striking exception of the large peak culminating to 1.6 $Å^2$ at position 46.

Even if, as emphasized by Artymiuk et al.,[16] effects other than true atomic motions (vibration) such as conformational substates, translational and rotational diffusion, and lattice disorder are not preponderant, they nevertheless contribute to the ms fluctuations. This might explain the fact that the ms fluctuations derived from B-factors are generally larger than the computed ms fluctuations.

With regard to the large peak appearing in the ms fluctuation curve computed by the normal mode analysis, the discrepancy between theoretical and experimental results is similar to the one occurring during the regularization step where loop 42–55 shows a large motion with respect to the X-ray coordinates. As has been discussed previously, it is likely that loop 42–55 is trapped in the crystal. This might explain the fact that the fluctuations deduced from the B-factors are smaller compared to the fluctuations calculated using the normal mode analysis. This also explains the absence of some of the peaks in area 118–125. In the crystal the residue Arg-119 of M2C4 is in close contact with the residues in the loop and therefore the fluctuations might propagate from the loop to this residue and, in turn, to the residues in the close vicinity.

Because HEWL[26] shows the same behavior as HL in area 44–52, we built also the crystal (tetragonal $a=b=79.1$ Å and $c=37.9$ Å, space group $P4_32_12$, one molecule in each of the 8 asymmetric units) and checked the environment of the loop in the same way as previously done for HL. The loop is close to three other molecules in the crystal involving, respectively, residues 125–129 (minimum distance: 2.03 Å) in the first one, residues 43–45 and 68 (minimum distance: 3.2 Å) in the second one, and residue 19 in the third one (minimum distance: 5.6 Å). The loop appears to be encaged within these fragments of

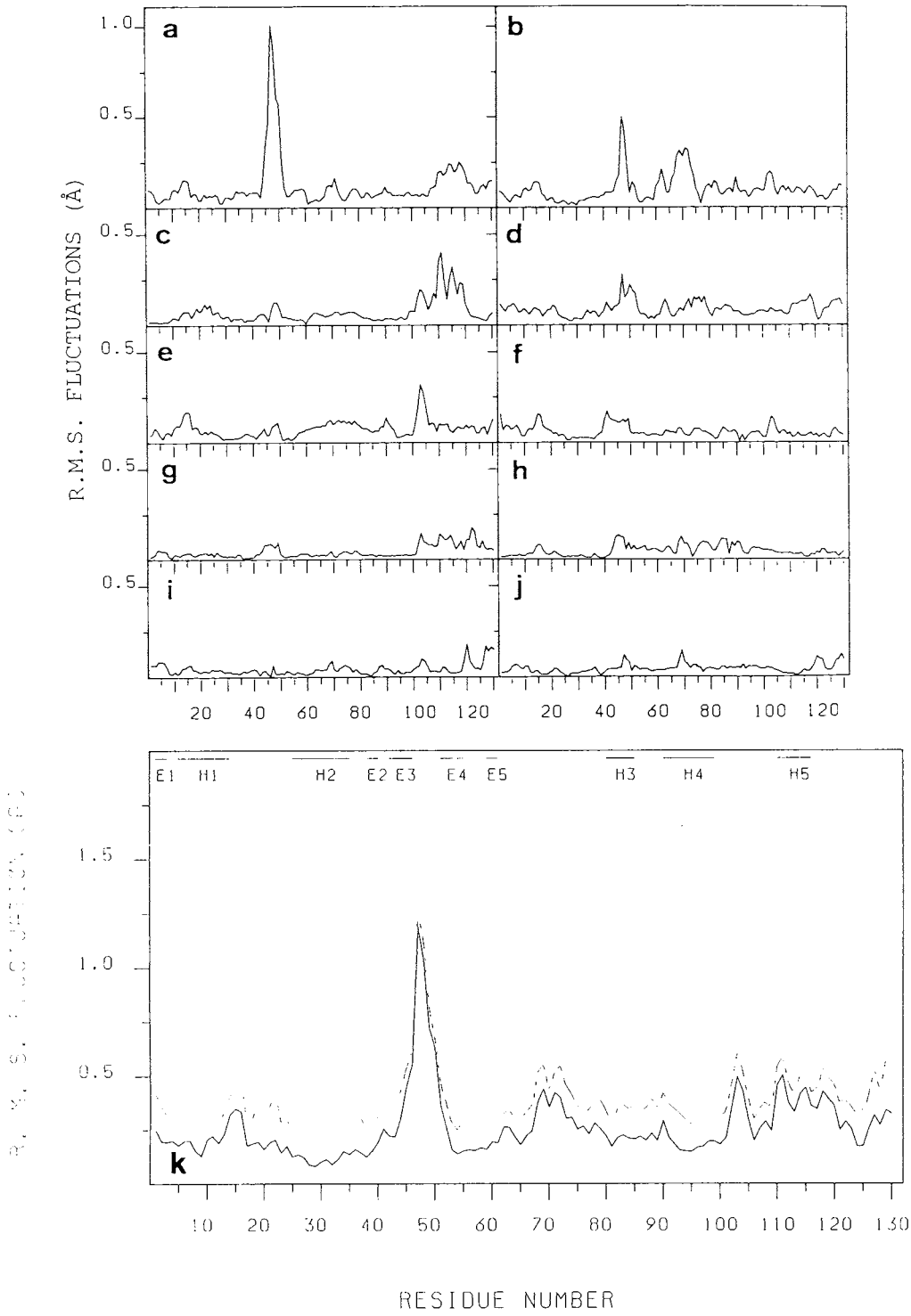


Fig. 8. Figures **a** to **j** give the rms fluctuations (Å) of the $C^{\alpha}$ position due to the 10 lowest normal modes considered individually. **k** gives the same plot for the sum of the 10 previous modes (solid line) as well as the rms fluctuations of the $C^{\alpha}$ position due to all the modes (broken line). In **k** the location of the secondary structures are indicated.

protein and might also be restrained in its motion. However, as we did not perform the minimization of HEWL it is difficult to judge precisely the motion undergone by the loop and to be absolutely positive about it.

It has to be noticed that the curve of mean fluctuations for HEWL (Figure 1a in Artymiuk et al.[16]) does not show a peak for residue 119 but shows a small peak at position 19. From the residues close to the loop in the crystal and according to the same

**TABLE IV. Fraction of the Mean Square Fluctuations Due to the First N Normal Modes Expressed in Percentage**

| Mode No. | % of ms fluctuation |
|---|---|
| 1 | 20 |
| 2 | 29 |
| 3 | 35 |
| 4 | 41 |
| 5 | 46 |
| 6 | 48 |
| 7 | 50 |
| 8 | 52 |
| 9 | 54 |
| 10 | 55 |

kind of argument used for HL, we can surmise that the motion of the loop in HEWL is somewhat trammeled by residue 19 (residues 43–45, which are closer, might also be involved, but this effect is difficult to bring out).

### Motion Due to Individual Normal Modes With Very Low Frequency

In the previous sections, we saw that the behavior of the dynamics is dominated by a few normal modes with very low frequency. In this section we are concerned with the motions of the protein due to these modes (only the first five modes are considered here). Figure 11 shows the main chain together with the displacement vectors for atoms of the main chain (N, $C^{\alpha}$, C′). The magnification of the displacement vectors is 8. Figure 12 shows another way of looking at the motions within the protein due to the lowest modes. These triangular maps show the change in the distance between all the possible pairs of residues (see method section).

1. Normal mode #1: frequency = 3.72 $cm^{-1}$. Figure 11a shows that the motion due to this mode concerns mainly the edges of the cleft (residues 45–51 in the second domain and residues 110–115 in the first domain). Indeed Figure 12a shows that residues 45–51 move with respect to all the other residues (the thin, broken band in the drawing). The map shows some motion inside area 45–51 itself. Residues 45–51 are going further apart from residues 90–130 as well as residues 1–40 (opening the cleft). They are going closer to residues 60–85. Relative motions in the area 45–55 itself are relatively erratic. The region 60–90 moves with respect to region 5–30 (getting closer) as well as area 105–125 (getting further apart mainly).

Triangles along the diagonal 1–40, 55–200, 105–130 show regions not affected by internal motions (rigid parts). Similarly the two regions 1–40 and 95–130, i.e., practically all the first domain, do not move relatively to each other in this mode.

2. Normal mode #2: frequency = 5.29 $cm^{-1}$. In Figure 11b the motions are mainly apparent in three regions 45–60, 66–73, 102–105. The motion of region 45–50 appears to close the cleft. The distance map of Figure 12b confirms this. Region 45–50 moves with respect to the rest of the molecule. Region 60–90 moves also with respect to 1–40 and 100–130. The two regions 1–40 and 100–130 do not show any relative motion. The three triangles along the diagonal 1–40, 55–95, 95–130 are still present even if some residues inside region 55–95 show some relative motion.

3. Normal mode #3: frequency = 6.17 $cm^{-1}$. In Figure 11c the fluctuations are mainly limited in region 108–120, some fluctuation acting to open the cleft, some having the opposite effect. Figure 12c shows a broad band indicating that region 100–120 is moving relatively to the rest of the molecule. This band is split into two parts. One part, roughly speaking, region 100–110, is getting closer to region 1–90. The other part, region 110–120, is getting further apart from region 30–120. The rest of the map with the exception of region 10–30, which shows some motion with respect to regions 45–50, and small regions around 60 and 70 do not show relative motion.

4. Normal mode #4: frequency = 6.41 $cm^{-1}$. Figure 11d shows small fluctuations involving mainly region 46–50. The resulting motion is an opening of the cleft. Figure 12d shows that only regions 1–30 and 120–130 do not move relatively to each other. According to the distance map, this mode does not show any distinct pattern.

5. Normal mode #5: frequency = 7.38 $cm^{-1}$. Figure 11e shows some fluctuations localized in region 100–110. Figure 12e shows two triangles along the diagonal 1–35 and 40–90. Region 100–110 shows some motion relatively to the rest of the molecule except with regions 30–45 and 90–100. Region 60–90 also shows motion with regions 95–130 and 5–35.

### Relative Motion of the Two Domains

We first started by examining the hinge bending motion due to the normal modes with very low frequency considered individually, i.e., normal mode #1: 3.72 $cm^{-1}$, normal mode #2: 5.29 $cm^{-1}$, normal mode #3: 6.17 $cm^{-1}$, and normal mode #4: 6.41 $cm^{-1}$.

The procedure for these four modes has been the same, first the change in all the distances between the heavy atoms of the two domains due to the normal mode in question has been calculated, then we searched for the parameters describing a hinge bending motion which give the best fit with these data. We started a series of minimizations from different points, trying to perform an extensive search in the parameters space. Usually the same set of starting points has been used for all the modes. In some cases, additional starting points were tried when we thought they could lead to a new minimum. However, for each mode we found a single

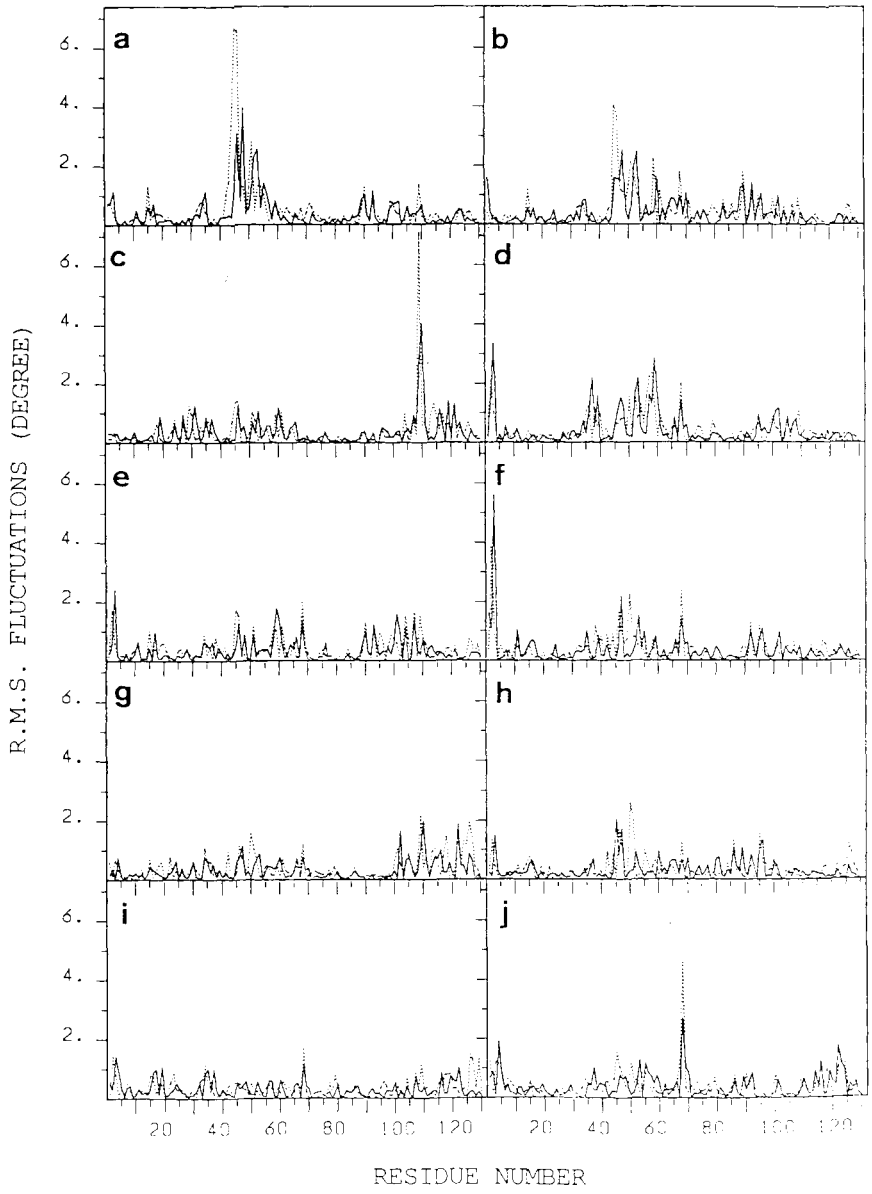


Fig. 9. Rms fluctuations of $\phi$ (solid line) and $\psi$ (dotted line) due to the 10 lowest frequency normal modes individually.

minimum, which of course is different for a different normal mode. Table V gives a summary of the results. Brooks and Karplus,[24] by the method described in the Methods section, found a normal mode with a frequency 3.6 $cm^{-1}$ having an overlap of 0.87 with the hinge bending mode determined in the work of McCammon et al.[1] This method did not allow them to characterize precisely the hinge bending motion but they estimated the hinge point to be located at Thr-40 and at the disulfide bond between Cys-76 and Cys-94. This is very close to the result we obtained for the first normal mode (3.7 $cm^{-1}$), with a hinge axis going roughly through the $C^{\alpha}$s 55 and 64. For the two first modes we found an axis which is located more or less between the two domains. For the third and fourth modes the axes found are quite excentered and the rms value for the rotation angle is rather small. It is likely that, for these modes, the partition of the molecule is not very good. Whatever the normal mode, the method is able to find a single minimum, but the characteristics of the hinge bending motion such defined are quite different. Therefore this raises the question of what is going to be the effect when the 10 lowest modes are used together for determining the change in the distances.

We thus determined the change in the distances due to the combination of the 10 modes with the

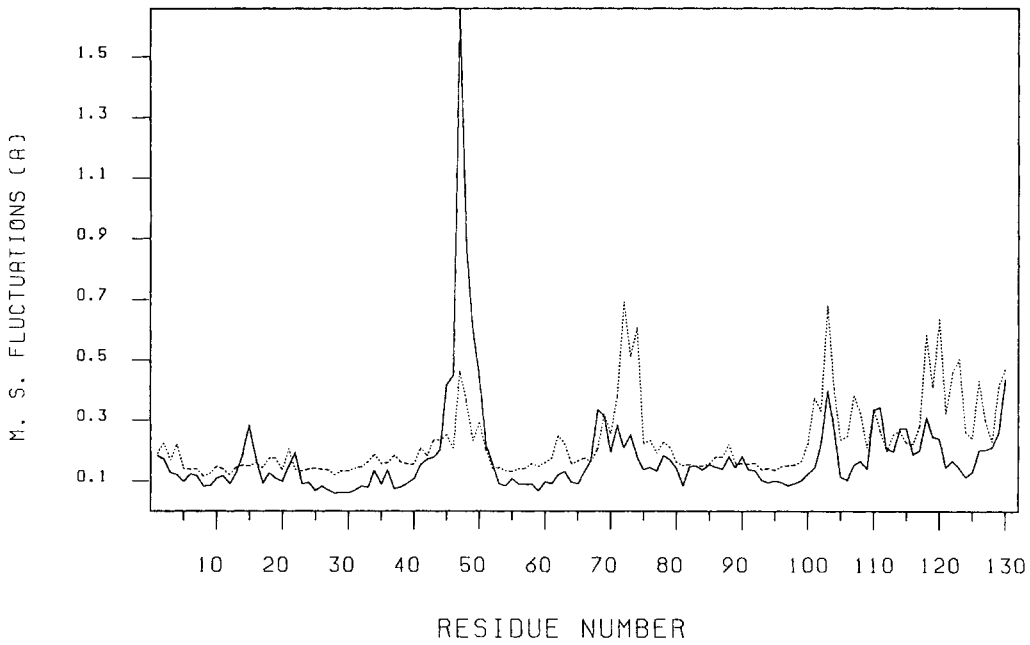


Fig. 10. Average mean square fluctuations (Å) of the position of the N, $C^{\alpha}$, $C^{\beta}$, C', O atoms for each residue. The solid line is the calculated curve and the dotted line is the curve deduced from the B-factors.

lowest frequency. As shown in Figure 8k these modes alone provide a good description of the protein dynamics. The data obtained were used in the same way described above. The same set of starting points was used and we added the position of the minima previously found to check if they remained as local minima in the new space to be explored. Nothing such was found, the minimization revealed only one minimum. At this point the value of the function is 1599 $Å^4$. The hinge axis goes through the $C^{\alpha}$s 55 and 76 and the rms value of the angle of rotation is 3.37°. This hinge axis is located, roughly, at the base of the β-sheet of the second domain and is very similar to the one found for describing the motion due to the first normal mode. The angle between these two axes is 7°. Apparently the contribution of the other modes results only in a slightly larger rms value of the angle of rotation without changing fundamentally the direction of the axis. For all practical purposes, the hinge bending motion can be described with a good approximation by the normal mode with the lowest frequency (3.72 $cm^{-1}$), i.e., it corresponds to the softest motion of the protein.

Although the method always finds a single minimum, this is not sufficient for assessing the agreement of the data with the model. A simple manner of doing this is to use the value of the residual at the minimum point. This residual is the sum of the squares of the difference between the points calculated using the model and the "experimental points." If the model is not used, i.e., if in Eq. (9) we set all the calculated distances to 0 (this corresponds to a rotation angle equal to 0) we obtain for the function the figures shown in the first column of Table V. The decrease of these initial values to the values at the minimum points gives an idea of the agreement of the model with the data. The figures in the third column of Table V show what percentage of the initial value the value of the function at the minimum points represents. The model gives a reasonable agreement with the data for the first normal mode (28%); for the other modes the agreement decreases, especially for the third mode. The result for the combination of the first 10 modes is similar to the result obtained for the first mode. This is consistent with the finding above that the first mode alone provides a good description of the hinge bending motion.

Another problem is to determine the parts of the molecule for which the model does not fit well the data. For each pair of residues $R_i$ and $R_j$ (the indices $i$ and $j$ refer, respectively, to the residues of the first and second domain) we define the following index:

$$\mathrm{Ind}(R_i, R_j) = \frac{\sum_a \sum_b \{[\Delta d_{ab}(\mathbf{x})]^2 - \langle(\Delta d_{ab})^2\rangle\}}{\sum_a \sum_b \langle(\Delta d_{ab})^2\rangle} \quad (10)$$

where $a$ and $b$ are indices for the heavy atoms of $R_i$ and $R_j$, respectively. This index is a signed quantity and represents for a pair of residues in each domain the relative error made in approximating $\langle(\Delta d_{ab})^2\rangle$ by $[\Delta d_{ab}(\mathbf{x})]^2$. The resulting plot (data not shown) is

**TABLE V. Results of the Least Square Minimization for the First Four Normal Modes Considered Individually (1–4) and for the Combination of the 10 First Modes (S)**

| | Initial* | Residual† | Percentage | $<(\Delta\theta)^2>^{1/2}$‡ | Axis§ | |
|---|---|---|---|---|---|---|
| 1 | 3495 | 963 | 28% | 2.88 | $C^{\alpha}55$ (0.82) | $C^{\alpha}64$ (0.92) |
| 2 | 269 | 152 | 45% | 1.12 | $C^{\alpha}36$ (0.33) | $C^{\alpha}93$ (1.48) |
| 3 | 201 | 143 | 71% | 0.70 | $C^{\alpha}10$ (1.92) | $C^{\alpha}17$ (1.49) |
| 4 | 25 | 13 | 52% | 0.94 | $C^{\alpha}109$ (1.86) | $C^{\alpha}112$ (1.52) |
| S | 6847 | 1599 | 23% | 3.37 | $C^{\alpha}55$ (0.58) | $C^{\alpha}76$ (0.39) |

*Value of the function (Å$^4$) if the angle of rotation is 0°.
†Value of the function at the minimum point (Å$^4$).
‡rms value of the angle of rotation in degree.
§The direction of the hinge axis is expressed by the two $C^{\alpha}$s whose position is the closest of the real axis. The value in parentheses gives the distance in Å between the corresponding $C^{\alpha}$ and the axis.

somewhat difficult to interpret, however a detailed examination reveals some rough features. It appears that the parts of the molecule for which the model gives the best agreement are both edges of the cleft. This is hardly surprising because the edges of the cleft are the areas in the protein for which the motions are important and therefore for which $\langle(\Delta d_{ab})^2\rangle$ are large. The method, in order to decrease significantly the target function, must take care of these regions. The model does not fit well the data, i.e., the change in the distances, between residues in helix C on the one hand and most of the residues of the second domain on the other hand. Similarly changes in the distances between many residues of the "bulky" part of the first domain (20–30 and 100–130) and residues located at the base of the loops in the second domain are not well represented.

In spite of these subtle discrepancies, the hinge bending motion appears as a whole as a reasonable model of the dynamics of human lysozyme, the first mode alone providing a good description of the hinge bending motion.

### Change Occurring in the Accessible Surface Area of Residues

The accessible surface area (asa), using only the heavy atoms, has been computed for the minimized structure and the thermally excited structures (i.e., the structures resulting of atomic motions due to the 10 lowest normal modes considered individually). As examples of the thermally excited structures, we consider, for each mode, two structures corresponding to

$$\sigma_i = \pm \langle \sigma_i^2 \rangle^{1/2} = \pm \frac{(k_B T)^{1/2}}{(2\pi\nu_i)} \qquad (11)$$

Table VI gives the accessible surface area for the whole molecule in each case (the radius of the probe, the water molecule, is $r = 1.4$ Å). A few points have to be noticed. First the variation in the accessible surface area is weak (−0.24 to +0.23%). Second the two first normal modes which are responsible for the large motions do not contribute significantly to the change in the accessible surface. Two modes are mainly responsible for the change in the accessibility surface area, modes 3 and 5. We paid special attention to the residues belonging to the active site cleft as the thermal motion of the protein is thought to have some biological significance. These residues appear as three clusters in the second domain: 35, 36, 37 then 44, 46, 51, 53, 57, 58, 59, 60, and 63, 64, 74, 75, and an (almost) continuous region in the first domain: 99, 100, 101, 102, 103, 104, 108, 109, 110, (114), 115. The average absolute value of asa variation for these residues is 0.1 Å$^2$ for the main chain and 0.5 Å$^2$ for the side chain. For all the residues in the protein the values are 0.1 Å$^2$ for the main chain and 0.4 Å$^2$ for the side chain. Therefore the residues in the active site cleft do not behave very differently of the other residues in the protein as far as the asa is concerned. Even though the normal modes with very low frequencies cause the largest motions in the protein, they do not affect drastically its accessible surface area. Increasing the radius of the probe first to 2.0 Å and then to 2.6 Å leads to results approximatively similar. The main difference is that the first and, to a lesser extent, the second and fifth normal modes are the ones which now contribute significantly to the change of accessible surface area. One could have thought that the opening of the cleft would result in a better exposition of the residues within the cleft, allowing thus a better interaction between the protein and its substrate. The results above appear to rule out such a naive concept. The motion, itself, might be only a device adopted by the protein to reach another substate (another local minimum in the neighborhood) where presumably the substrate–protein interaction is easier.

### Change in the Topology of Atom Packing

Figure 11 shows spatial continuity of the displacement vectors. The motions appear as fluid-like, that is the magnitude and orientation of the displacement vectors are only slightly different between a given atom and its neighbors. In order to study these motions in more detail we considered the change in the topology of atom packing in the molecule when

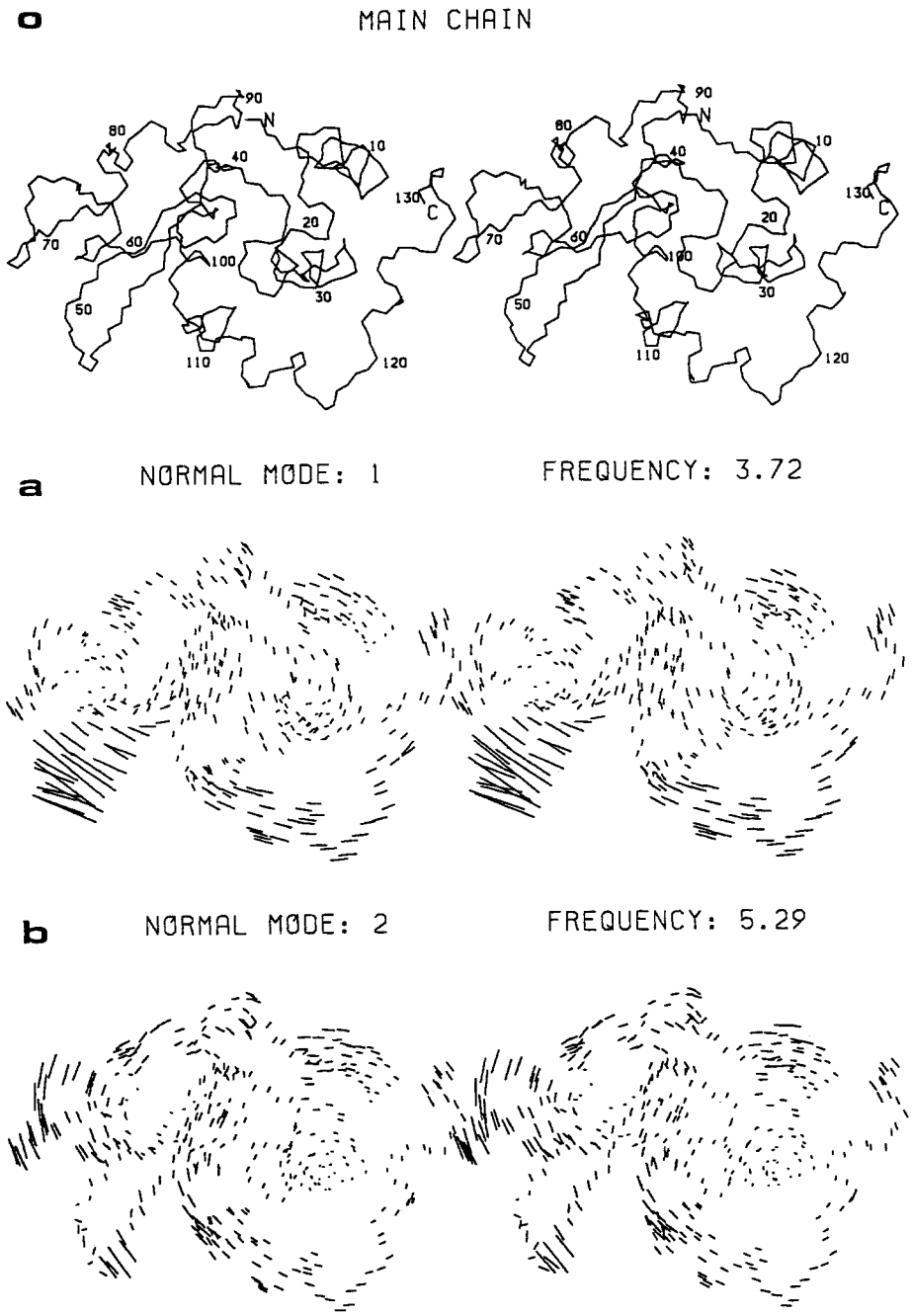


Fig. 11. Stereo drawings of the molecule (o) together with the displacement vectors due to the five lowest frequency normal modes (**a**–**e**). The magnification of these vectors is 8. Figs. **c**, **d**, **e** are on page 276.

the atoms undergo thermal fluctuations. The change in the topology of atom packing has been defined by Noguti and Gō[18] in their study of the multiple minima populating the conformational energy surface. They showed that the local deformations observed between a pair of minima are usually accompanied by a rearrangement in atom packing. Here, we are interested in characterizing the thermally excited state, that is, to examine if the thermal fluctuations can generate a change in the topology of atom packing. As far as the normal modes are considered individually, with the exception of 4 pairs in the excited structure resulting from atomic fluctuations due to the first normal mode, there is no change in the topology of atom packing in the molecule. Two types of pairs can be examined: (1) the pairs of atoms which are not in contact in both structures (the minimized one and the thermally ex-

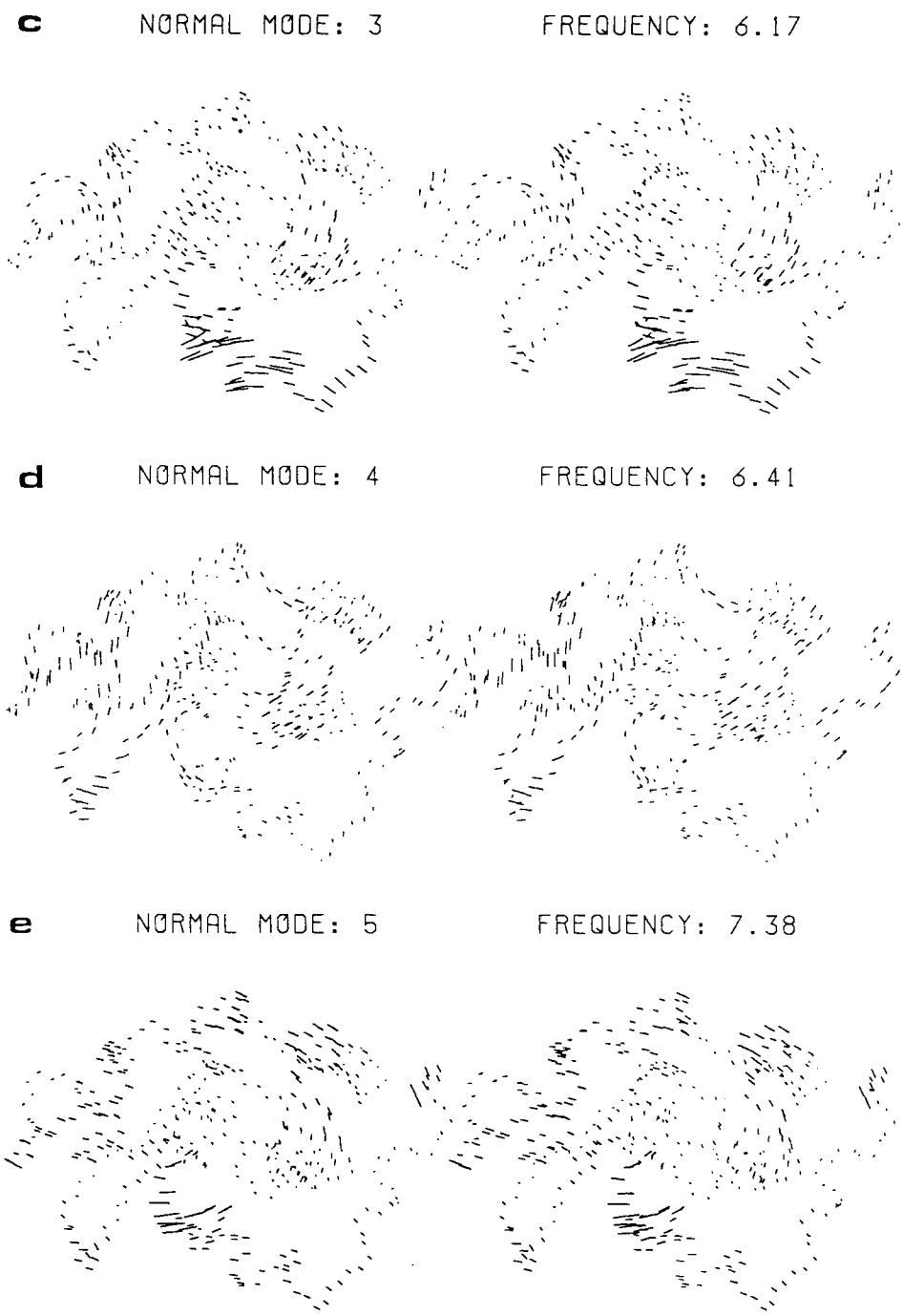


Fig. 11, c-e. Legend appears on page 275.

cited one) and (2) the pairs of atoms which are in contact at least in one of these two structures. This latter set, which contains about 25,000 pairs for human lysozyme, is used to assess the change occurring in the packing of atoms. The maximum value for the change in the distance between a pair of atoms for the two sets described above is, respectively, 1.55 Å and 0.56 Å for the first normal mode, 0.94 Å and 0.40 Å for the second normal mode, 0.68 Å and 0.48 Å for the third normal mode, and 0.45 Å and 0.25 Å for the fourth normal mode. The fluctuations of the distances between pairs of atoms in contact are therefore much smaller than between pairs of distant atoms in the protein. This is consistent with the picture of a fluid-like motion given by Figures 11. If the 10 lowest normal modes are collectively

**TABLE VI. Total Accessible Surface Area (Å²) for the Protein**

| Normal mode number | $\sigma_i$ | | |
|---|---|---|---|
| | $+ \langle \sigma_i^2 \rangle^{1/2}$ | 0 | $- \langle \sigma_i^2 \rangle^{1/2}$ |
| 1 | 6380.7 | 6378.3* | 6381.5 |
| 2 | 6378.4 | 6378.3 | 6382.8 |
| 3 | 6372.8 | 6378.3 | 6392.9 |
| 4 | 6381.4 | 6378.3 | 6376.6 |
| 5 | 6362.8 | 6378.3 | 6392.9 |
| 6 | 6378.3 | 6378.3 | 6381.0 |
| 7 | 6378.3 | 6378.3 | 6378.7 |
| 8 | 6378.1 | 6378.3 | 6378.7 |
| 9 | 6376.5 | 6378.3 | 6382.9 |
| 10 | 6378.8 | 6378.3 | 6381.0 |

*Value for the minimum energy structure.

taken into account, there are 34 pairs in the molecule for which the distance changes by more than 0.5 Å. In that case, the maximum value for the change of the distance among pairs in contact (second set) is 1.04 Å. The change in the topology of atom packing involves four regions of the protein: the first loop in the second domain (residue 45 on one side and residues 46, 47, and 50 on the other side), an interface between the two loops of the second domain (pairs involving residues 63–49 and 87–41), an interface between the first and the second domain in the active site cleft (pair involving residues 64–102), and part of the active site cleft belonging to the first domain (residues 111, 112, 115 on one side and residues 27 and 34 on the other side). As could be expected, the change in the topology of atom packing involves residues belonging to the edges of the cleft. As previously shown, the edges of the active site are the part which undergo the largest motions in the molecule. However, the number of pairs concerned is small and, thus, under the assumption of harmonicity of the potential energy, the motion can be described as very nearly elastic.

## DISCUSSION AND CONCLUSION

The main assumption of the normal mode analysis is that the potential energy can be approximated by a quadratic function of the generalized coordinates in the vicinity of the minimum. Gō et al.[10] have shown that, along directions of the normal modes, the energy curves generally increase more rapidly than the assumed parabolic behavior. This fact is particularly prominent for the normal modes with low frequency. It is therefore interesting to compare the results obtained by the normal mode analysis with the available experimental data. A comparison of the mean square fluctuations of the backbone atoms deduced from B-factors and obtained with the normal mode analysis reveals that the agreement in term of position of the peak is, in the light of the previous observations, surprisingly good. The comparison of the amplitudes is reasonable though it is more difficult to interpret because the ms fluctuations deduced from the B-factors include terms other than vibrations (such as lattice disorder, multiple substates, translation, and rotation) and that there are reasons to believe that the motion of the loop is somewhat hampered by the packing in the crystal. On the other hand, the large peak observed at position 45 in the $C^\alpha$ rms fluctuations is certainly exaggerated in the normal mode analysis.

The shape of the ms fluctuations of the $C^\alpha$ can be explained to a large extent by several normal modes with very low frequency, each contributing a pattern or part of a pattern (Fig. 8). The rest of the normal modes contribute to what we termed "background" fluctuations, i.e., apparently patternless fluctuations. This raises the question whether there are some features characteristic of the normal modes with very low frequencies. According to Figure 7, it appears that each mode among the 400 first ones involves about 325 heavy atoms, i.e., roughly one-third of the protein (the first mode stands as an exception involving only 200 heavy atoms, one fifth of the molecule.) The proportion of $C^\alpha$ atoms, backbone atoms, and side chain atoms is nearly the same between these modes and very close to the proportion existing in the protein. Therefore, the modes with higher frequency (up to the 400th normal mode) are not really different from the modes with very low frequency. The only difference is that the amplitude for each mode, as previously mentioned, depends on the inverse of the square of the frequency and thus decreases rapidly. Therefore the modes with very low frequency impose their specific patterns on the dynamics of the molecule while the patterns of other modes contribute to the "background" fluctuations. These patterns, i.e., the position of the peaks depends critically on the eigenvectors. The good agreement in the position of the peaks between experimental and computed data means that the method is able to find such subspace in the dihedral angle space which corresponds to the modes with very low frequencies. This is important because it shows that we can be quite confident in the description of the dynamics in term of a few collective motions of parts of the protein, even if the amplitudes of such motions are more difficult to assess precisely.

In this article we propose a method for determining the hinge motion which best fits the data obtained from the normal mode analysis. This method does not presuppose any prior knowledge of the position of the hinge axis. The hinge axis we found is indeed notably different from the hinge axis previously postulated and the rms value of the angle of rotation is twice as large as the value previously calculated. This latter fact shows that the hinge bending motion is softer than previously thought and therefore results in larger fluctuations. A detailed observation of the results shows that the

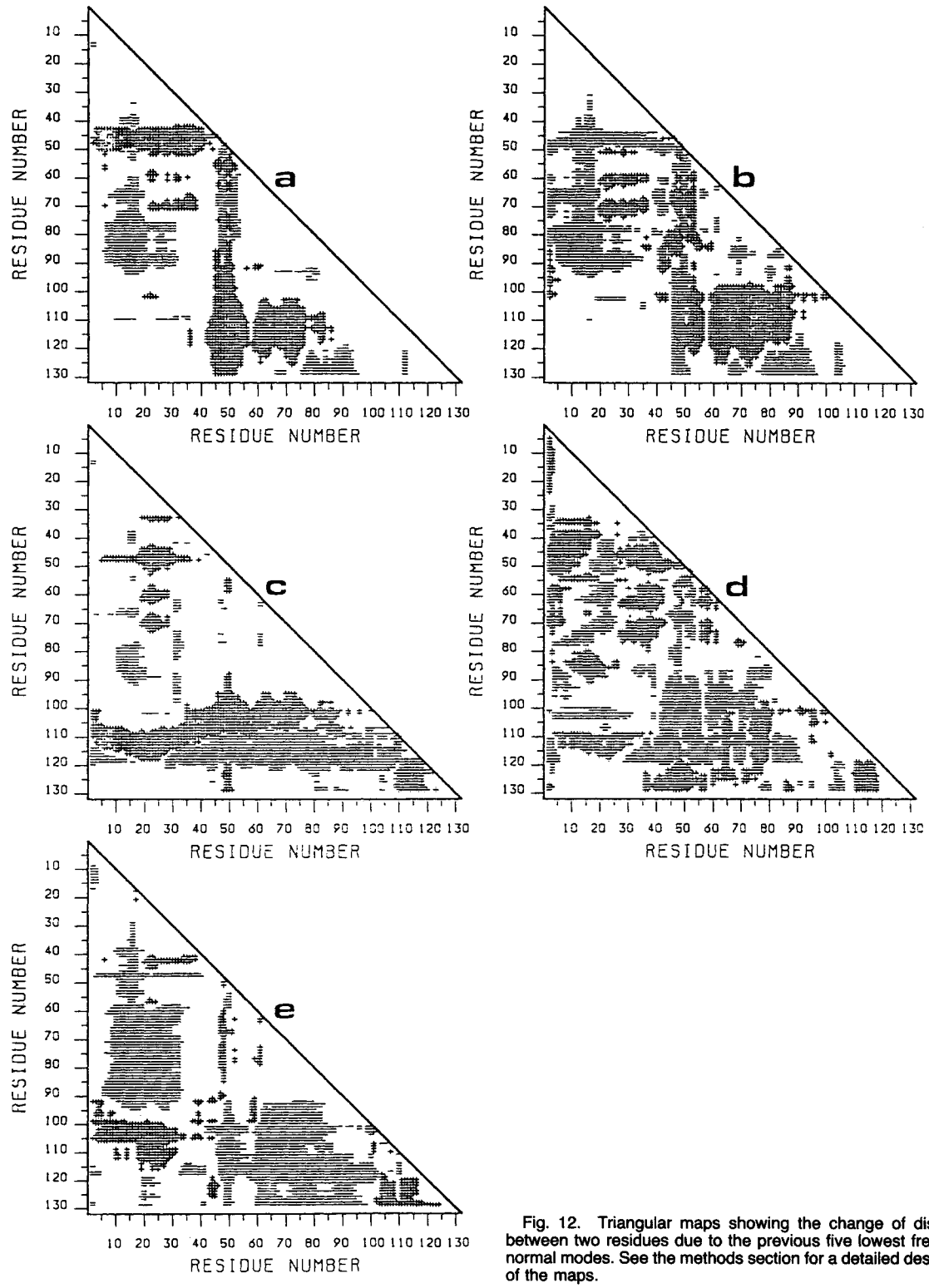


Fig. 12. Triangular maps showing the change of distances between two residues due to the previous five lowest frequency normal modes. See the methods section for a detailed description of the maps.

hinge bending motion provide a fairly good description of the dynamics of human lysozyme.

Surprisingly enough, the observed motions do not result in a better exposure of the residues within the cleft to the solvent. The change in the accessible surface area for these residues is weak and similar to the change in the accessible surface area of other, less mobile, residues in the protein.

A study of the topology of atom packing shows that there is little change in the environment of a given atom during the motion. Therefore the change in the orientation and module of displacement vectors for atoms in contact is more or less monotonic, that is, within a small region there is no abrupt change in the magnitude and, maybe more important, in the direction of the displacement vectors. This is characteristic of an elastic deformation.[18] This might be a consequence of the uniqueness of the minimum in the conformational space. The system undergoes some deformations but it always fluctuates around a single equilibrium position. In reality, the system can pass over energy barriers within the range of thermal fluctuations and is able to reach new minima. In this latter case the changes in the magnitude and especially the direction of the displacement vectors are likely to be important, leading to plastic deformations. Clearly more work is needed for the characterisation of the thermally excited state, which takes into account the anharmonicity of the potential energy surface. A study is now in progress which considers the change in the topology of atom packing between thermally excited conformations along the trajectory of a Monte Carlo simulation (in dihedral angle space).

In many systems it is thought that, at least, two different conformations with different functionalities exist which are connected by a low energy pathway. The normal mode analysis cannot give a direct picture of this phenomenon, but can provide some clue about the pathway of low energy which is followed by the protein during the transition. This pathway is likely to correspond to the subspace spanned by the normal modes with very low frequency. In this respect, in the case of proteins for which the transition between two different conformations is triggered by a cofactor, it would be interesting to see the effect of the binding of the cofactor upon the normal modes.

## ACKNOWLEDGMENTS

We thank Dr. T. Noguti and Dr. J. Higo for helpful discussions. J-F.G. is the recipient of a J.S.P.S. postdoctoral fellowship. Computation has been done at computer centers of Kyoto University and of the Institute for Molecular Science. This work has been supported by grants-in-aid from the Ministry of Education, Science and Culture, Japan.

## REFERENCES

1. McCammon, J.A., Gelin, B., Karplus, M., Wolynes, P.G. The hinge-bending mode in lysozyme. Nature (London) 262:325–326, 1976.
2. Huber, R., Bennett, W.S. Functional significance of flexibility in proteins. Biopolymers 22:261–279, 1983.
3. Weber, I.T., Steitz, T.A. Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution. J. Mol. Biol. 198:311–326, 1987.
4. Remington, S.J., Wiegand, G., Huber, R. Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 Å resolution. J. Mol. Biol. 158:111–152, 1982.
5. Karplus, M., McCammon, J.A. The internal dynamics of globular proteins. CRC Crit. Rev. Biochem. 9:293–349, 1981.
6. Levitt, M. Protein conformation, dynamics and folding by computer simulation. Annu. Rev. Biophys. Bioeng. 11: 251–271, 1982.
7. Van Gunsteren, W.F., Berendsen, H.J.C., Hermans, J., Hol, W.G.J., Postma, J.P.M. Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray. Proc. Natl. Acad. Sci. U.S.A. 80:4315–4319, 1983.
8. Levy, R.M., Keepers, J.W. Computer simulation of protein dynamics: Theory and experiment. Comments Mol. Cell. Biophys., 3:273–294, 1986.
9. Noguti, T., Gō, N. Efficient Monte Carlo method for simulation of fluctuating conformations of native proteins. Biopolymers 24:527–546, 1985.
10. Gō, N., Noguti, T., Nishikawa, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc. Natl. Acad. Sci. U.S.A. 80:3696–3700, 1983.
11. Brooks, B., Karplus, M. Harmonic dynamics of proteins: Normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc. Natl. Acad. Sci. U.S.A. 80:6571–6575, 1983.
12. Levitt, M., Sander, C., Stern, P.S. The normal modes of a protein: Native bovine pancreatic trypsin inhibitor. Int. J. Quant. Chem.: Quant. Biol. Symp. 10:181–199, 1983.
13. Levitt, M., Sander, C., Sterm, P.S. Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. J. Mol. Biol. 181:423–447, 1985.
14. Nishikawa, T., Gō, N. Normal modes of vibration in bovine pancreatic trypsin inhibitor and its mechanical property. Proteins 2:308–329, 1987.
15. Artymiuk, P.J., Blake, C.C.F. Refinement of human lysozyme at 1.5 Å resolution: Analysis of non-bonded and hydrogen-bond interactions. J. Mol. Biol. 152:737–762, 1981.
16. Artymiuk, P.J., Blake, C.C.F., Grace, D.E.P., Oatley, S.J., Phillips, D.C., Sternberg, M.J.E. Crystallographic studies of the dynamic properties of lysozyme. Nature (London) 280:563–568, 1979.
17. Gō, N. A theorem on amplitudes of thermal atomic fluctuations in large molecules assuming specific conformations calculated by normal mode analysis. Biophys. Chem. 35:105–112, 1990.
18. Noguti, T., Gō, N. Structural basis of hierarchical multiple substates of a protein. IV: Rearrangements in atom packing and local deformations. Proteins 5:125–131, 1989.
19. Némethy, G., Pottle, M.S., Scheraga, H.A. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. J. Phys. Chem. 87:1883–1887, 1983.
20. Noguti, T., Gō, N. A method of rapid calculation of a second derivative matrix of conformational energy for large molecules. J. Phys. Soc. Jpn. 52:3685–3690, 1983.
21. Abe, H., Braun, W., Noguti, T., Gō, N. Rapid calculation of first and second derivatives of conformational energy with respect to dihedral angles for proteins. General recurrent equations. Comp. Chem. 8:239–247, 1984.
22. Wako, H., Gō, N. Algorithm for rapid calculation of hessian of conformational energy function of proteins by supercomputer. J. Comp. Chem., 8:625–635, (1987).
23. Noguti, T., Gō, N. Dynamics of native globular proteins in terms of dihedral angles. J. Phys. Soc. Jpn. 52:3283–3288, 1983.
24. Brooks, B., Karplus, M. Normal modes for specific motions of macromolecules: Application to the hinge-bending mode of lysozyme. Proc. Natl. Acad. Sci. U.S.A. 82:4995–4999, 1985.
25. Sternberg, M.J.E., Grace, D.E.P., Phillips, D.C. Dynamic information from protein crystallography: An analysis of temperature factors from refinement of the hen-egg-white-lysozyme structure. J. Mol. Biol. 130:231–253, 1979.
26. Blake, C.C.F., Koenig, D.F., Mair, G.A., North, A.C.T., Phillips, D.C., Sarma, V.P. Structure of hen egg-white lysozyme. Nature (London) 206:757–761, 1965.