Journal Title: J Mol Biol
ISSN: 0022-2836
Article Author(s): Levitt M
Article Title: Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme.
Year: 1985 Feb 5
Volume: 181
Issue: 3
Pages: 423-47
PMID: 2580101
User's Comments: In color, if available

J. Mol. Biol. (1985) 181, 423–447

Protein Normal-mode Dynamics: Trypsin Inhibitor, Crambin, Ribonuclease and Lysozyme

Michael Levitt, Christian Sander† and Peter S. Stern

Chemical Physics Department
Weizmann Institute of Science
76100 Rehovot, Israel

(Received 2 November 1983, and in revised form 25 September 1984)

We have developed a new method for modelling protein dynamics using normal-mode analysis in internal co-ordinates. This method, normal-mode dynamics, is particularly well suited for modelling collective motion, makes possible direct visualization of biologically interesting modes, and is complementary to the more time-consuming simulation of molecular dynamics trajectories.

The essential assumption and limitation of normal-mode analysis is that the molecular potential energy varies quadratically. Our study starts with energy minimization of the X-ray co-ordinates with respect to the single-bond torsion angles. The main technical task is the calculation of second derivative matrices of kinetic and potential energy with respect to the torsion angle co-ordinates. These enter into a generalized eigenvalue problem, and the final eigenvalues and eigenvectors provide a complete description of the motion in the basic 0·1 to 10 picosecond range. Thermodynamic averages of amplitudes, fluctuations and correlations can be calculated efficiently using analytical formulae.

The general method presented here is applied to four proteins, trypsin inhibitor, crambin, ribonuclease and lysozyme. When the resulting atomic motion is visualized by computer graphics, it is clear that the motion of each protein is collective with all atoms participating in each mode. The slow modes, with frequencies of below 10 cm^{-1} (a period of 3 ps), are the most interesting in that the motion in these modes is segmental. The root-mean-square atomic fluctuations, which are dominated by a few slow modes, agree well with experimental temperature factors (B values). The normal-mode dynamics of these four proteins have many features in common, although in the larger molecules, lysozyme and ribonuclease, there is low frequency domain motion about the active site.

1. Introduction

Internal dynamics, especially collective motion, is important for protein function (see review by Huber & Bennett, 1983). By collective motion we mean a process in which a substantial part of the protein moves as a unit relative to other parts. A recent beautiful example provided by X-ray crystallography is citrate synthase (Remington *et al.*, 1982; Chothia & Lesk, 1984). In each catalytic cycle, the large and small domain of the enzyme first close around the substrate for catalytic action, and then open a cleft between them for product release and the next round of substrate binding. A number of other proteins also have (at least) two different conformations in two functionally distinct states. The transition from one to the other must be a major mode of internal motion, i.e. there must be a low-energy pathway of collective motion connecting the two, made possible by the overall three-dimensional structure. Neither the stages along this pathway nor the time-scale of the transition have yet been observed experimentally. It is clear, however, that such processes, which involve the correlated motion of large masses, must be slower than vibrations involving only localized atomic motion.

Experimental study of protein dynamics by spectroscopic observation of low-frequency vibrations has recently become possible (see review by Peticolas, 1979). For example, the neutron time-of-flight spectrum of lysozyme below 100 cm^{-1} shows a broad peak at 75 cm^{-1} and a shoulder at 25 cm^{-1} (Bartunik *et al.*, 1982); a new Raman technique (Genzel *et al.*, 1976) also gives peaks at these two frequencies. Similar Raman peaks are seen in α-chymotrypsin at about 29 cm^{-1} (Brown *et al.*, 1972) and in acid phosphatase at about 25, 50

† Present address: Department of Biophysics, Max Planck Institute of Medical Research, Jahnstrasse 29, 6900 Heidelberg, West Germany.

0022-2836/85/030423-25 $03.00/0

and 80 cm^{-1} (Twardowski, 1978). In some cases the observation of these peaks is dependent on conformation and/or solvent environment, suggesting that they represent collective internal vibrations involving major parts of the protein. The exact nature of the motion responsible for the low-frequency Raman and neutron spectroscopic peaks is still unknown.

The theoretical study of protein dynamics has involved the calculation and analysis of molecular dynamics trajectories (Karplus & McCammon, 1981; van Gunsteren *et al.*, 1983; Levitt, 1983*a*,*b*). The atomic positions and velocities as a function of time are given by numerical solution of the classical equations of motion using detailed interatomic potentials and starting from the X-ray co-ordinates. In principle, molecular dynamics provides a very powerful tool for simulating realistic protein motion, especially when water molecules are included. In practice, however, there are two major drawbacks: (1) the elementary time step is so small that simulations longer than 100 ps become prohibitively costly on present computers; (2) the analysis of the thousands of co-ordinate sets along the trajectory is tedious and subject to statistical uncertainty. Consequently, study of slow processes, such as collective motion or conformational rearrangements, is difficult.

Normal-mode dynamics of proteins offer an alternative to the simulation of molecular dynamics trajectories and are ideally suited to the study of slow collective motion. The normal modes are coupled vibrations. They are found by assuming that the molecular potential energy can be approximated as a quadratic or harmonic function of the dynamic variables and then solving a generalized eigenvalue problem to give a closed analytical description of the motion. The eigenvalues give the vibrational time-scales (frequencies) and the eigenvectors give the details of the corresponding motion. The motion can be described and visualized at each separate frequency or as a more complicated superposition of modes. Time-averaged equilibrium and kinetic properties of the system, such as positional fluctuations or time correlation functions, can be calculated accurately and efficiently as weighted sums. The price paid for all this is the limited accuracy of the harmonic approximation and the difficulty associated with the inclusion of water molecules. The faults and virtues of the two methods are complementary in that molecular dynamics provide an approximate numerical solution to the exact equations of motion, whereas normal-mode dynamics provide an exact analytical solution to the approximate equations of motion.

Normal-mode or vibrational analysis is well established and much used in molecular physics (Wilson *et al.*, 1955): it is very successful in reproducing vibrational spectra of many small molecules (Shimanouchi, 1970) and of homo-biopolymers (Itoh & Shimanouchi, 1970). The quadratic approximation to the potential energy is obtained by a Taylor expansion of the tota molecular potential energy about its minimu using atomic Cartesian co-ordinates (Lifson Warshel, 1968) or internal co-ordinates (bon lengths, bond angles and torsion angles) as th dynamic variables. Application to large non repetitive molecules is technically difficult. The firs protein normal-mode calculation (Tasumi *et al* 1982) on glucagon (29 residues) uses a specia α-helical force field without non-bonded term approximates side-chains as point masses and doe not minimize energy. Noguti & Gō (1982) analyse collective fluctuations of trypsin inhibitor (5 residues) in terms of 241 torsion angle co-ordinate by diagonalizing the potential energy secon derivative matrix. In a subsequent paper, Gō *et a* (1983) present a more complete treatment involvin simultaneous diagonalization of both kinetic an potential energy second derivative matrices; thi work and our preliminary study of trypsin inhibito (Levitt *et al.*, 1983) are the first normal-mod studies of a protein in atomic detail. More recently Brooks & Karplus (1983) used 1740 Cartesian co ordinates in a complete normal-mode analysis o this same small protein.

In the present study, we have developed completely general method that can be applied t calculate the normal-mode dynamics of any syste of point masses in terms of any set of generalize dynamic variables. Harmonic approximations o both the potential and kinetic energy matrices ar obtained by a combination of convergent energy minimization, numerical differentiation, and the us of a best-superposition algorithm.

We test the method on a hexapeptide usin single-bond torsion angles as the degrees of freedo and obtain results identical to those of the standar Cartesian co-ordinate treatment.

We use the method to calculate the torsion angl normal modes of four proteins: trypsin inhibito (Huber *et al.*, 1971; Deisenhofer & Steigemann 1975), crambin (Hendrickson & Teeter, 1981) ribonuclease (Richards & Wyckoff, 1971; Wlodawe & Sjölin, 1983) and lysozyme (Blake *et al.*, 1965 1967). Although these proteins are of different size and have different chain foldings, many features o their normal-mode dynamics are similar. (1) Th lowest frequency for each protein is betwee 2 cm^{-1} and 5 cm^{-1}, while the average frequency close to 60 cm^{-1} for all. The contribution of eac mode to the atomic motion decreases rapidly wit increasing frequency. Very few of the mode (between 1 and 3%) are responsible for most (70% of the atomic motion of the backbone. These low frequency modes are characterized by rigid-bod motion of α-helices and smooth bending an twisting deformation of β-sheets. (2) The averag root-mean-square fluctuation of α-carbon atoms about 0·6 Å, while that of ϕ and ψ angles is abou 16°. Fluctuations of the backbone atoms a markedly reduced in regions of hydrogen-bonde secondary structure and are larger in loops. (3) Th agreement between temperature factors calculate

from normal modes and derived from the crystallographic study is good, with a correlation coefficient between 0·54 and 0·65.

Most attention is focused on the analysis of the normal-mode dynamics of bovine pancreatic trypsin inhibitor as this protein has been studied very extensively by experimental and theoretical methods. Calculations are performed using two BPTI† conformations, the refined X-ray co-ordinates and a set of co-ordinates obtained at the end of a 130 ps dynamics trajectory (Levitt, 1983*b*). The analysis of BPTI showed that: (1) the distribution of frequencies (density of states) and variation of α-carbon atom fluctuations along the chain are very similar in the two conformations, which differ by 1·8 Å r.m.s. (α-carbon atoms only). (2) The motion as visualized on a graphics display involves bending, twisting, arm-swinging and breathing of segments about 10 Å in extent (4 to 5 residues). (3) The different regions of secondary structure have characteristic frequencies: β-sheet from 5 to 30 cm^{-1} α-helix from 50 to 110 cm^{-1} and turns from 5 to 75 cm^{-1}. (4) Torsion angles generally move in an uncorrelated fashion except when the residues involved are near-neighbours along the chain or connected by hydrogen bonds.

Normal-mode dynamics can be used as an additional tool for the study of the dynamic behaviour of biological macromolecules, complementing X-ray crystallography, spectroscopy and molecular dynamics simulations. Their application to trypsin inhibitor, crambin, ribonuclease and lysozyme affords us a new view of protein dynamics.

2. Methods

(a) *General theory of normal modes*

(i) *Equations of motion*

The theory of normal modes (Goldstein, 1950) gives a complete analytical solution to the equations of motion subject to the assumption that the potential energy of the system can be approximated as a *quadratic* function of the n co-ordinates, q_i, in the vicinity of the potential energy minimum at q_i^0:

$$E_P = \tfrac{1}{2}\sum_{i,j}^{n} F_{ij}(q_i-q_i^0)(q_j-q_j^0) = \tfrac{1}{2}(\mathbf{q}-\mathbf{q}^0)^T\mathbf{F}(\mathbf{q}-\mathbf{q}^0). \quad (1)$$

The kinetic energy is also approximated as a *quadratic* function of the velocities $\dot{q}_i(=dq_i/dt)$:

$$E_K = \tfrac{1}{2}\sum_{i,j}^{n} H_{ij}\dot{q}_i\dot{q}_j = \tfrac{1}{2}\dot{\mathbf{q}}^T\mathbf{H}\dot{\mathbf{q}}. \quad (2)$$

The co-ordinates $\mathbf{q}$ are generalized and can be atomic Cartesian co-ordinates, rigid-body translations and rotations, torsion angles or any other dynamic variables chosen to describe the essential degrees of freedom. The equations of motion in terms of any set of co-ordinates are given by Lagrange's equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \left(\frac{\partial L}{\partial q_i}\right),$$

† Abbreviations used: BPTI, bovine pancreatic trypsin inhibitor; r.m.s., root-mean-square.

where $L = E_K - E_P$. From eqns (1) and (2) we have:

$$\partial L/\partial\dot{q}_i = \sum_j^n H_{ij}\dot{q}_j \quad \text{and} \quad \partial L/\partial q_i = -\sum_j^n F_{ij}(q_j-q_j^0)$$

and Lagrange's equations become:

$$\sum_j^n H_{ij}\ddot{q}_j = -\sum_j^n F_{ij}(q_j-q_j^0). \quad (3)$$

These equations are solved by assuming that the solution is the vector $\mathbf{q}$ with elements q_j of the form:

$$q_j = q_j^0 + \sum_k^n A_{jk}\alpha_k \cos(\omega_k t+\delta_k). \quad (4)$$

This gives:

$$\ddot{q}_j = \mathrm{d}^2q_j/\mathrm{d}t^2 = -\sum_k^n A_{jk}\alpha_k\omega_k^2 \cos(\omega_k t+\delta_k).$$

Substituting for q_j and $\ddot{q}_j$ in eqn (3) gives:

$$-\sum_j^n H_{ij}\sum_k^n A_{jk}\alpha_k\omega_k^2\cos(\omega_k t+\delta_k) = -\sum_j^n F_{ij}\sum_k^n A_{jk}\alpha_k\cos(\omega_k t+\delta_k)$$

or

$$\sum_k^n\left(\sum_j^n H_{ij}A_{jk}\right)\alpha_k\omega_k^2\cos(\omega_k t+\delta_k) = \sum_k^n\left(\sum_j^n F_{ij}A_{jk}\right)\alpha_k\cos(\omega_k t+\delta_k). \quad (5)$$

Because eqn (5) must hold for all times, t:

$$\left(\sum_j^n H_{ij}A_{jk}\right)\omega_k^2 = \sum_j^n F_{ij}A_{jk} \quad (6)$$

for $k = 1$ to n. Rewriting eqn (6) in matrix notation gives the equations of motion as:

$$\mathbf{HA\Lambda} = \mathbf{FA}, \quad (7)$$

where $\mathbf{\Lambda}$ is diagonal and $\Lambda_{ii} = \omega_i^2$, i.e. the frequencies of vibration of the system.

Before eqn (7) can be solved to give $\mathbf{A}$ and $\mathbf{\Lambda}$, it is necessary to set a normalization condition. Defining $Q_k = \alpha_k\cos(\omega_k t+\delta_k)$ gives from eqn (4):

$$q_j - q_j^0 = \sum_k^n A_{jk}Q_k \quad \text{or} \quad \mathbf{q}-\mathbf{q}^0 = \mathbf{AQ} \quad (8)$$

as the relation between the generalized co-ordinates $\mathbf{q}$ and the co-ordinates $\mathbf{Q}$. Expressing the potential and kinetic energies in terms of $\mathbf{Q}$ gives:

$$E_P = \tfrac{1}{2}(\mathbf{q}-\mathbf{q}^0)^T\mathbf{F}(\mathbf{q}-\mathbf{q}^0) = \tfrac{1}{2}\mathbf{Q}^T\mathbf{A}^T\mathbf{FAQ}$$

$$E_K = \tfrac{1}{2}\dot{\mathbf{q}}\mathbf{H}\dot{\mathbf{q}} = \tfrac{1}{2}\dot{\mathbf{Q}}^T\mathbf{A}^T\mathbf{HA}\dot{\mathbf{Q}}. \quad (9)$$

If the normalization condition is chosen so that $\mathbf{A}^T\mathbf{HA} = \mathbf{1}$, this gives:

$$E_K = \tfrac{1}{2}\dot{\mathbf{Q}}^T\dot{\mathbf{Q}} = \tfrac{1}{2}\sum_k^n \dot{Q}_k^2.$$

This also means that by multiplying the equations of motion (eqn (7)), on the left by $\mathbf{A}^T$:

$$\mathbf{A}^T\mathbf{FA} = \mathbf{A}^T\mathbf{HA\Lambda} = \mathbf{\Lambda},$$

which gives:

$$E_P = \tfrac{1}{2}\mathbf{Q}^T\mathbf{\Lambda Q} = \tfrac{1}{2}\sum_k^n \Lambda_{kk}Q_k^2 = \tfrac{1}{2}\sum_k^n \omega_k^2Q_k^2. \quad (9a)$$

Because the potential and kinetic energies are now simple sums of squares of Q_i and $\dot{Q}_i$, the co-ordinates $\mathbf{Q}$ are normal co-ordinates. Because

$$E_k = \tfrac{1}{2}\sum_k^n \dot{Q}_k^2$$

involves no coefficients, the co-ordinates **Q** are also mass-scaled.

Solution of eqn (7) for **A** and **Λ** with the normalization condition $\mathbf{A}^T\mathbf{H}\mathbf{A} = \mathbf{1}$ is accomplished by standard methods, for example, subroutine **F02AEF** from the Numerical Algorithms Group Library (Wilkinson & Reinsch, 1971). The complete dynamic behaviour of the system is then given by:

$$q_j = q_j^0 + \sum_k^n A_{jk}\alpha_k \cos(\omega_k t + \delta_k), \tag{10}$$

where α_k is the amplitude, ω_k the angular frequency and δ_k the phase of the kth normal-mode of motion. The angular frequency is given by $\omega_k = \Lambda_{kk}^{\frac{1}{2}}$ and the phase, δ_k, and amplitude, α_k, depend on the positions and velocities at time $t = 0$.

Before this formalism can be applied to any system, values are required for $\mathbf{q}^0$, **F** and **H**. The equilibrium co-ordinates, $\mathbf{q}^0$, for which the system has a minimum potential energy can be found by numerical methods used to minimize general functions, for example, **VA09D** from the Harwell Subroutine Library (Fletcher, 1972). For efficient minimization, values of the gradient $\partial E_P/\partial q_j$ should be calculated analytically. The matrix **F**, which contains the second derivatives of the potential energy with respect to the generalized co-ordinates, **q**, is given by $F_{ij} = \partial^2 E_P/\partial q_i\,\partial q_j$ at the equilibrium position $\mathbf{q}^0$. **F** can be calculated analytically, but it is often easier to use numerical differentiation of $\partial E_P/\partial q_i$ to give:

$$F_{ij} = \frac{1}{\varepsilon}\left[\left(\frac{\partial E_P}{\partial q_i}\right)_{q_j = q_j^0 + \varepsilon} - \left(\frac{\partial E_P}{\partial q_i}\right)_{q_j = q_j^0}\right]. \tag{11}$$

The matrix **H**, which contains the second derivatives of the kinetic energy with respect to the generalized velocities $\dot{\mathbf{q}}$, is given by $H_{ij} = \partial^2 E_K/\partial\dot{q}_i\partial\dot{q}_j$ at the equilibrium position $\mathbf{q}^0$. If the system consists of N points with masses m_l and Cartesian co-ordinates $\mathbf{r}_l$, the kinetic energy is given by:

$$E_K = \tfrac{1}{2}\sum_l^N m_l \dot{\mathbf{r}}_l^2. \tag{12}$$

Small changes in the generalized co-ordinates, δq_i, cause small changes in the Cartesian co-ordinates that are given by:

$$\delta\mathbf{r}_l = \sum_i^n (\partial\mathbf{r}_l/\partial q_i)\,\delta q_i.$$

If these changes occur in time δt, we have:

$$\delta\mathbf{r}_l/\delta t = \sum_i^n (\partial\mathbf{r}_l/\partial q_i)\,\delta q_i/\delta t$$

or

$$\dot{\mathbf{r}}_l = \sum_i^n (\partial\mathbf{r}_l/\partial q_i)\dot{q}_i.$$

Substituting for $\dot{\mathbf{r}}_l$ in eqn (12) gives:

$$E_K = \tfrac{1}{2}\sum_l^N m_l \sum_i^n \frac{\partial\mathbf{r}_l}{\partial q_i}\dot{q}_i \sum_j^n \frac{\partial\mathbf{r}_l}{\partial q_j}\dot{q}_j$$

$$= \tfrac{1}{2}\sum_i^n\sum_j^n \left(\sum_l^N m_l \frac{\partial\mathbf{r}_l}{\partial q_i}\cdot\frac{\partial\mathbf{r}_l}{\partial q_j}\right)\dot{q}_i\dot{q}_j.$$

Comparison with eqn (2) gives:

$$H_{ij} = \sum_l^N m_l \frac{\partial\mathbf{r}_l}{\partial q_i}\cdot\frac{\partial\mathbf{r}_l}{\partial q_j}. \tag{13}$$

Special attention must be paid to the calculation of $\partial\mathbf{r}_l/\partial q_i$. Because the kinetic energy must not include overall translation and rotation of the system, the change $\delta\mathbf{r}_l$ caused by δq_i must not change any other **q** co-ordinate and not move or rotate the centre of mass. A completely general numerical method of the calculation of **H** is given in section (b), below. For $m_l = 1$, the matrix **H** is the metric of the co-ordinate space; a distance $|\Delta\mathbf{r}|^2$ in Cartesian space is related to the distance in **q** space by

$$|\Delta\mathbf{r}|^2 = \sum_{i,j}^n \Delta q_i H_{ij} \Delta q_j.$$

This formalism can be used to solve the equations of motion of any system of point masses in any arbitrary system of generalized co-ordinates. The steps to be followed are: (1) define the Cartesian co-ordinates, $\mathbf{r}_l$, in terms of the generalized co-ordinates, q_i, and minimize the potential energy with respect to the co-ordinates; (2) at the minimum, calculate **F** and **H** using eqns (11) and (13); (3) solve the equations of motion (eqn (7)) for **A** and **Λ** and express the motion of the generalized co-ordinates, **q**, in terms of the normal co-ordinates **Q** (eqns (8) and (10)).

(ii) *Time-averaged properties*

The particular path of atomic motion will depend on the initial conditions that determine the amplitude, α_k, and phase, δ_k, of each normal mode. Time-averaged properties of the motion depend only on the amplitudes α_k and are, therefore, of more general importance. Consider, for example, the correlation coefficient of pairs of co-ordinates that is given by $\langle\Delta q_i(\tau)\Delta q_j(\tau)\rangle$, where $\langle\ \rangle$ denote averaging over all time, τ. From eqn (8):

$$\Delta q_i(\tau) = q_i(\tau) - q_i^0 = \sum_k^n A_{ik} Q_k(\tau)$$

and

$$\langle\Delta q_i(\tau)\Delta q_j(\tau)\rangle = \left\langle \sum_k^n A_{ik}Q_k(\tau) \sum_l^n A_{jl}Q_l(\tau)\right\rangle$$

$$= \sum_k^n\sum_l^n A_{ik}A_{jl}\langle Q_k(\tau)Q_l(\tau)\rangle.$$

Now

$$Q_k(\tau) = \alpha_k\cos(\omega_k\tau + \delta_k) \quad \text{and} \quad Q_l(\tau) = \alpha_l\cos(\omega_l\tau + \delta_l),$$

so that:

$$\langle Q_k(\tau)Q_l(\tau)\rangle = \begin{cases} 0, & \text{if } k \neq l \\ \alpha_k^2/2, & \text{if } k = l \end{cases}$$

(derived from

$$\frac{1}{\tau_0}\int_0^{\tau_0} \cos^2(\omega\tau + \delta)\,\mathrm{d}\tau = 1/2 \text{ as } \tau_0 \to \infty).$$

Thus:

$$\langle\Delta q_i(\tau)\Delta q_j(\tau)\rangle = \tfrac{1}{2}\sum_k^n A_{ik}A_{jk}\alpha_k^2. \tag{14}$$

Generalization of the correlation coefficient to values of co-ordinates that are at time interval t apart gives:

$$\langle\Delta q_i(\tau + t)\Delta q_j(\tau)\rangle = \tfrac{1}{2}\sum_k^n A_{ik}A_{jk}\alpha_k^2\cos\omega_k t \tag{15}$$

(derived as above using

$$\langle Q_k(\tau + t)Q_l(\tau)\rangle$$

$$= \frac{1}{\tau_0}\int_0^{\tau_0} \alpha_k\cos(\omega_k(\tau + t) + \delta_k)\alpha_l\cos(\omega_l\tau + \delta_l)\,\mathrm{d}\tau$$

$$= \begin{cases} 0, & \text{if } k \neq l; \\ \tfrac{1}{2}\alpha_k^2\cos\omega_k t, & \text{if } k = l). \end{cases}$$

Finally, consider some parameter Δp_i of the system that is a linear function of the change in the generalized co-ordinates Δq_k, i.e.:

$$\Delta p_i = \sum_k^n P_{ik}\Delta q_k.$$

The correlation coefficients of the parameters Δp_i are given by:

$$\begin{aligned}\langle \Delta p_i(\tau)\Delta p_j(\tau)\rangle &= \sum_l^n \sum_m^n P_{il}P_{jm}\langle \Delta q_l(\tau)\Delta q_m(\tau)\rangle \\ &= \tfrac{1}{2}\sum_k^n \sum_l^n \sum_m^n P_{il}P_{jm}A_{lk}A_{mk}\alpha_k^2 \qquad (16)\\ &= \tfrac{1}{2}\sum_k^n P'_{ik}P'_{jk}\alpha_k^2,\end{aligned}$$

where

$$P'_{ik} = \sum_l^n P_{il}A_{lk} \quad \text{and} \quad P'_{jk} = \sum_m^n P_{jm}A_{mk}.$$

One of the most commonly used transformations is from generalized co-ordinates **q** to Cartesian co-ordinates **r**. In this case $P_{il} = \partial \mathbf{r}_i/\partial q_l$; the transformation matrix **P** is also used in the calculation of **H**.

(iii) *Thermal amplitudes*

The amplitude α_k of a particular normal mode depends on the temperature. In classical dynamics, each normal mode will have a time-averaged potential energy of $\frac{1}{2}k_BT$ above the value at the minimum (k_B is Boltzmann's constant and T is the absolute temperature). The potential energy of any one mode is $\frac{1}{2}\omega_k^2 Q_k^2(\tau)$ (eqn (9a)) and the time-averaged value is:

$$\begin{aligned}\tfrac{1}{2}\omega_k^2\langle Q_k^2(\tau)\rangle &= \tfrac{1}{2}\omega_k^2\alpha_k^2\langle \cos^2(\omega_k\tau+\delta_k)\rangle \\ &= \tfrac{1}{4}\omega_k^2\alpha_k^2 = \tfrac{1}{2}k_BT.\end{aligned} \qquad (17)$$

Thus, the classical mean-square fluctuation of Q_k is:

$$\langle Q_k^2(\tau)\rangle_{cl} = k_BT/\omega_k^2.$$

This also gives:

$$\alpha_k = (2k_BT/\omega_k^2)^{1/2},$$

so that all the time-averaged properties dependent on α_k also depend on temperature.

In quantum mechanical dynamics, each mode will behave like an harmonic oscillator with energy levels $\hbar\omega_k$ apart, where $\hbar$ is Planck's constant/2π. The mean-square fluctuation of the normal-mode variable Q_k is related to the temperature (Kubo, 1967) by:

$$\langle Q_k^2(\tau)\rangle_{qm} = \frac{\hbar}{2\omega_k}\coth\left(\frac{\hbar\omega_k}{2k_BT}\right) = \tfrac{1}{2}\alpha_k^2. \qquad (18)$$

When y is small,

$$\coth(y) = (e^y+e^{-y})/(e^y-e^{-y})$$

tends to $1/y$ and $\langle Q_k^2(\tau)\rangle_{qm}$ tends to k_BT/ω_k^2, which is the classical value. At 300 K, $y = \nu_k/417$ for the frequency ν_k ($=\omega_k/2\pi$) expressed in reciprocal centimetres. Because ν_k is much less than 200 cm^{-1} for most of the protein normal modes calculated here, y is small and the quantum mechanical and classical treatments give almost identical results at room temperature.

(b) *Normal modes of a protein*

The above formalism is completely general. Here it is applied to the calculation of the normal modes of 4 proteins: BPTI, crambin, ribonuclease A and lysozyme. All co-ordinates were taken from the Brookhaven Protein Data Bank and have reference numbers 4PTI, 1CRN, 4RSA and 6LYZ, respectively.

The generalized co-ordinates used are the ϕ, ψ and χ torsion angles about single bonds, together with the 6 co-ordinates of the rigid-body translation and rotation of the entire molecule. The number of residues and torsion angles are as follows: BPTI, 58 residues, 208 angles; crambin, 46 residues, 139 angles; ribonuclease, 124 residues, 455 angles; and lysozyme, 129 residues, 471 angles. The potential energy function, the scheme for calculating first derivatives and the minimization algorithm are those used in a previous study of protein folding (Levitt, 1983*c*) and have been described therein. The S–S bonds are kept closed by a bonding force, but no other restraints are imposed. It is important to note that the minimum energy conformation must be determined precisely for the quadratic expansion of the potential energy (eqn (1)) to hold. This is achieved by use of double-precision arithmetic (16 significant figures) and the convergent **VA09D** minimization algorithm (Fletcher, 1972; see Harwell Subroutine Library).

Once the equilibrium values of the torsion angles, q_j^0, have been determined, the matrix elements $F_{ij} = \partial^2 E_P/\partial q_i\,\partial q_j$ are calculated by numerical differentiation (eqn (11)) with $\varepsilon = 10^{-9}$ radians. This gives F_{ij} values accurate to 8 decimal places. For BPTI, minimization of the potential energy to convergence requires less than 500 evaluations of the energy and its first derivatives and calculation of the **F** matrix requires another 208 evaluations. In principle, F_{ij} could have been calculated analytically using formulae given before (Levitt, 1972; Sander & Stern, 1979; Katz *et al.*, 1979; Noguti & Gō, 1983*b*). For a molecule as large as a protein, the analytical calculation is not very efficient; as **F** is only needed at one conformation, $\mathbf{q}^0$, numerical differentiation is quite satisfactory.

The matrix **H** is also calculated numerically as described above (eqn (13)). At the equilibrium co-ordinates $\mathbf{q}^0$, one of the torsion angle variables q_i is changed to $q_i = q_i^0+\varepsilon_i$, causing the Cartesian co-ordinates of the atoms to change from $\mathbf{r}_l^0$ to $\mathbf{r}_l$. Torsional rotation about a particular bond can move the centre of mass. This centre of mass motion must be eliminated by superimposing the co-ordinates **r** as a rigid body onto the initial co-ordinates $\mathbf{r}^0$. Use of transformation matrices (Kabsch, 1976; McLachlan, 1979), calculated with Cartesian co-ordinates multiplied by $\sqrt{m}$, gives a new set of Cartesian co-ordinates $\mathbf{r}'$ from which all overall rotation or translation of the molecule has been eliminated. The transformation $\partial \mathbf{r}_l/\partial q_i$ is now calculated as $(\mathbf{r}'_l-\mathbf{r}_l^0)/\varepsilon_i$. Repeating the procedure for every torsion angle gives the complete **H** matrix (see eqn (13)). Here $\varepsilon_i = 0{\cdot}001$ radians is found to work well. The transformation matrix $\partial \mathbf{r}_l/\partial \mathbf{q}_i$ also converts torsion angles to Cartesian co-ordinates and is used to calculate molecular properties from the normal modes (see section (a), (ii), above). An elegant analytical calculation of **H** has been published by Noguti & Gō (1983*a*).

(c) *Units*

The units used in these calculations require some explanation. Energy is expressed in kilocalories per mole, mass in grammes per mole and distance in ångstrom units. With these choices, 1 unit of time is equal to $t_0 = 4{\cdot}8888\times10^{-14}$ s (based on the U.S. National Bureau of Standards definition of 1 kcal $= 4{\cdot}184\times10^{10}$ erg ($1\ \text{erg} = 10^{-7}$ J)). An angular frequency of ω radians per time-unit is ω/t_0 radians per second and corresponds to a frequency $\nu = \omega/2\pi = \omega/(2\pi t_0)$ cycles per second. The frequency in reciprocal centimetres is $\nu = \omega/(2\pi c t_0)$ $= 108{\cdot}59\omega$ cm^{-1} (where c, the speed of light $= 2{\cdot}9979\times10^{10}$ cm/s).

(d) *Computing considerations*

One of the major advantages of protein normal-mode dynamics is that requirements of computer resources are

much smaller than with molecular dynamics. For the protein BPTI, which has 515 atoms and 208 single-bond torsion angles, one evaluation of the potential energy and its analytical first derivatives takes 20 s (all times are c.p.u. time for the VAX 11/780 computer). Reaching the potential energy minimum takes 2 h, calculating the **F** and **H** matrices numerically takes 90 min, and doing the full analysis presented here takes 30 min. Thus, the entire normal-mode treatment takes 4 h. For a protein twice as large as BPTI (e.g. lysozyme), the time requirement is 26 h (i.e. about 6 times as much). Virtual memory requirements are Z Mbyte for BPTI and 4 Mbytes for lysozyme.

3. Results

(a) *Tests on alanine hexapeptide*

The method introduced here for calculating the vibrational modes with respect to the single-bond torsion angles involves numerical differentiation and complicated matrix algebra. We felt it was essential to test the new method against results obtained with the established method for calculating the normal modes in Cartesian co-ordinates (Lifson & Warshel, 1968; Lifson & Stern, 1982). Using only torsion angle co-ordinates is equivalent to keeping all bond lengths and bond angles fixed. As the bond lengths and bond angles become stiffer, the low-frequency vibrations in Cartesian co-ordinates should become like those calculated with the subset of single-bond torsion angles. Test calculations done on the extended form of alanine hexapeptide show this trend. For very stiff bond lengths, bond angles and double-bond torsion angles, the vibrational frequencies and normal modes are identical in both co-ordinate systems. This severe test of the method using two completely independent programs verifies that the new method works.

(b) *Tests on four proteins*

(i) *Minimization in torsion angle space*

Normal-mode dynamics was begun by mini mizing the molecular potential energy with respec to the ϕ, ψ and χ single-bond torsion angles. Th starting conformations used were: for BPTI, X, th X-ray co-ordinates (Huber *et al.*, 1971; Deisenhofe & Steigemann, 1975) and E130, the co-ordinate after 130 picoseconds of molecular dynamics an minimization in Cartesian space (Levitt, 1983*b*); fo crambin, X, the X-ray co-ordinates (Hendrickson & Teeter, 1981); for ribonuclease, X, the X-ray co ordinates (Wlodawer & Sjölin, 1983) and EX, th co-ordinates after 100 steps of minimization i Cartesian space; for lysozyme, X, the X-ray co ordinates (Blake *et al.*, 1965,1967). The energ value after minimization, the r.m.s. deviation fro the X-ray structure and the number of energ evaluations are given in Table 1. More cycles o minimization are needed for the larger proteins; i all cases the energy change over the last 50 steps less than 0·1 kcal/mol and the final r.m.s. gradien is less than 10^{-5} kcal/mol radian.

The total potential energy (torsion plus non bond) of each protein is approximately dependen on the size of the molecule: the potential energy pe residue is between −4·3 and −5·8 kcal/mol. Th vibrational free energy ($\Delta H - T\Delta S$) is a mor accurate linear function of the size of the molecul the free energy per torsion angle is $0{\cdot}89 \pm 0{\cdot}02$ kca mol. The entropy per mode is also very similar fo all four proteins ($2{\cdot}50 \pm 0{\cdot}03\ k_B$).

Minimization causes a r.m.s. deviation from th X-ray structure of between 0·68 and 2·58 Å. Thes deviations are comparable to those found b molecular dynamics simulations (Levitt, 1983*a*). B studying two conformations of BPTI (TX an

Table 1

Minimum energy values and r.m.s. deviations

Protein		Residues	No. of torsion angles	Atoms	Energy (kcal/mol)				r.m.s. deviation Δr_A (Å)§	Evaluations to minimum
					Torsion	Non-bond	ΔH‡	$-T\Delta S$‡		
BPTI†	TX	58	208	515	14	−263	125	−322	1·54	520
	TE130	58	208	515	16	−293	125	−314	2·58	341
Crambin	TX	46	139	374	3	−206	84	−210	0·68	243
Ribonuclease	TX	124	455	1109	33	−729	274	−676	1·14	958
	TEX	124	455	1109	28	−747	274	−677	1·35	775
Lysozyme	TX	129	471	1167	32	−634	284	−695	0·89	687

† For BPTI, TX is the conformation obtained by torsion angle minimization from X, the X-ray co-ordinates, and TE130 is th corresponding conformation from E130, the co-ordinates after 130 ps of molecular dynamics at 300 K (Levitt, 1983*a*). For ribonuclease TEX is the conformation obtained by torsion angle minimization from EX, the co-ordinates after 100 steps of Cartesian co-ordina minimization from X, the X-ray conformation. For the other two proteins, the results presented are for the TX conformation.

‡ The vibrational enthalpy, ΔH, and entropy, ΔS, are calculated from the normal-mode frequencies ν_i using the established formul (Hagler *et al.*, 1979):

$$\Delta H = \sum_i \{h\nu_i/2 + h\nu_i/(e^{h\nu_i/k_BT} - 1)\}$$

$$-T\Delta S = \sum_i \{k_BT \ln (1 - e^{-h\nu_i/k_BT}) - h\nu_i/(e^{h\nu_i/k_BT} - 1)\},$$

where h is Planck's constant, T is the absolute temperature, and k_B is Boltzmann's constant. At $T = 300$ K, $k_BT/h = 417$ cm^{-1}.

§ The r.m.s. deviation is calculated after superposition of the co-ordinates as described before (Levitt, 1983*a*). All atoms are includ in the superposition and the calculation of the r.m.s. deviation.

TE130) and ribonuclease (TX and TEX), we test the sensitivity of normal-mode dynamics to shifts of this magnitude.

(ii) *Distribution of frequencies*

The frequency distributions of the four proteins are compared in Figure 1 and properties of these distributions are summarized in Table 2. These distributions are similar in that they are all skewed towards high frequency, the mean frequency is close to 60 cm^{-1} and the percentage of low frequencies (below 20 cm^{-1}) is about 15%. In each protein, there are a few frequencies greater than 220 cm^{-1}; these are due to the disulphide bonds that are the only bonds free to vibrate with torsion angle variables. (There are 3 S–S bonds in BPTI and crambin and 4 S–S bonds in ribonuclease and lysozyme.)

Other features of the distributions depend on the particular protein. The larger proteins, ribonuclease and lysozyme, have a smaller lowest frequency. The position of the peak frequency (see Table 2) indicates that crambin is the least flexible molecule and that BPTI and ribonuclease are most flexible. Relaxing the X-ray conformation in Cartesian space as done for BPTI (TE130) and ribonuclease (TEX) reduces the lowest frequency, but gives fewer modes below 20 cm^{-1}.

(iii) *Amplitude of vibration*

Table 2 also compares the average r.m.s. fluctuations of α-carbon atoms and backbone torsion angles. The ϕ and ψ fluctuations are very similar for the four proteins (mean $\sigma_\phi = 16{\cdot}2 \pm 0{\cdot}5°$, mean $\sigma_\psi = 16{\cdot}5 \pm 0{\cdot}5°$), with ψ consistently showing a slightly larger extent of motion. The α-carbon fluctuations show more variation (mean $\sigma_\alpha = 0{\cdot}56 \pm 0{\cdot}06$ Å). For BPTI the conformation obtained after 130 picoseconds of molecular dynamics is better packed, with a lower non-bond energy, and less flexible, with a lower vibrational entropy and smaller σ_α.

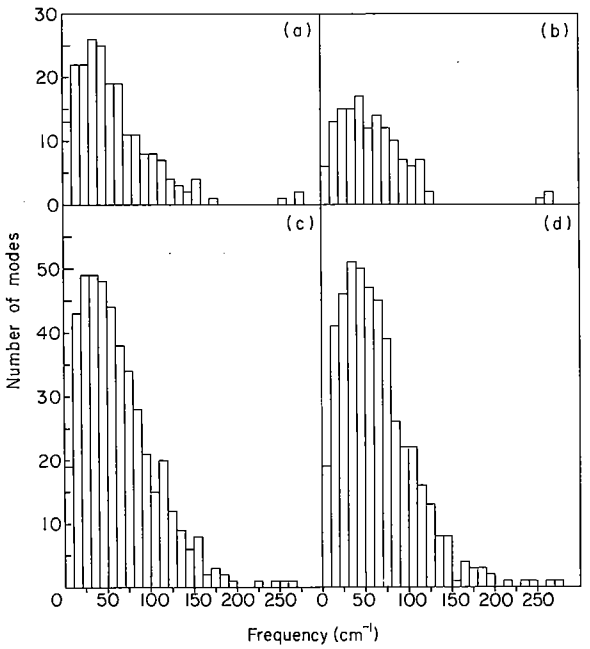


Figure 1. Showing the vibrational spectra calculated here with single-bond torsion angle variables. The number of modes with frequencies in a 10 cm^{-1} interval is plotted against frequency to give a density of states distribution. If each mode had the same intensity and a half line-width of 5 cm^{-1} the intensity envelope would be like the distribution shown. The spectra shown are for: (a) BPTI, (b) crambin, (c) ribonuclease and (d) lysozyme.

(c) *Motion in secondary structure*

The four proteins studied here have different tertiary structures, but each has regions of α-helix and β-sheet secondary structure. Here we show that these regions vibrate with distinctive frequencies and amplitudes. Because these findings were similar for the different proteins, detailed results are presented only for one protein, BPTI.

Figure 2 shows that the dominant frequencies of ϕ and ψ torsion angle vibration of BPTI depend on

Table 2

Comparison of frequencies and fluctuations

Protein		Frequency (cm^{-1})			Number of modes			r.m.s. fluctuations‡		
		Lowest	Peak	Mean	Total	<20 cm^{-1}	70% of σ_α†	σ_α(Å)	σ_ϕ(°)	σ_ψ(°)
BPTI	TX	4·6	38	58·7	208	35	8	0·55	16·6	16·9
	TE130	3·9	37	61·6	208	28	5	0·51	17·2	17·3
Crambin	TX	4·4	41	60·0	139	19	3	0·53	15·7	16·5
Ribonuclease	TX	2·3	34	62·9	455	58	4	0·67	15·7	15·8
	TEX	2·4	32	63·2	455	62	4	0·68	16·2	16·2
Lysozyme	TX	3·0	37	64·0	471	60	8	0·56	15·9	16·2

† The number given is m, where the average r.m.s. fluctuations of α-carbon atoms due to the m lowest modes, $\sigma_\alpha^{1\text{to}m}$, is just greater than or equal to $0{\cdot}7\sigma_\alpha$, i.e.:

$$\sigma_\alpha^{1\text{to}m} = \left\{ \sum_{k=1}^{m} (\sigma_\alpha^k)^2 \right\}^{1/2} \geq 0{\cdot}7\sigma_\alpha.$$

‡ σ_α is the average r.m.s. fluctuation of α-carbon atoms due to all modes at 300 K; σ_ϕ and σ_ψ are the corresponding average r.m.s. fluctuations of the ϕ and ψ torsion angles.

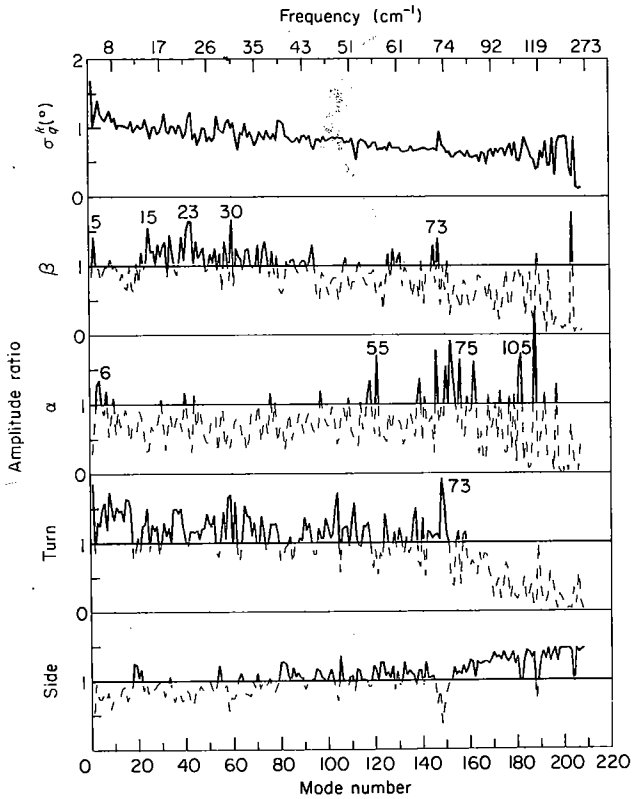


Figure 2. Showing how the average r.m.s. fluctuation of all 208 BPTI ϕ, ψ and χ torsion angles, σ_q^k, depends on frequency and secondary structure. The top graph plots σ_q^k averaged over all residues. The other curves plot the r.m.s. fluctuation averaged over only those torsion angles (ϕ, ψ, χ) in particular regions of the structure (β-sheet, α-helix, turns and side-chains). These values are normalized by dividing by σ_q^k to give an amplitude ratio defined as $\{\sum_i(\sigma_{qi}^k)^2\}^{1/2}/\sigma_q^k$, where the r.m.s. fluctuation of the ith torsion angle in mode k is $\sigma_{qi}^k = \langle q_i^k(\tau)^2\rangle^{1/2} = A_{ik}\alpha_k/2^{1/2}$ (from eqn (14)), and the sum is over all torsion angles in a particular class. The regions of secondary structure are as follows: β-sheet, 16 to 24, 29 to 36, 44 to 46; α-helix, 2 to 6, 47 to 56; turn, the remainder. In this paper, when we refer to α-helix we generally include 3_{10}-helix (e.g. residues 2 to 6 of BPTI). In the plots of the normalized regional r.m.s. fluctuations, values less than unity were de-emphasized by using light broken lines.

the nature of the secondary structure of the particular residue with which the torsion angle is associated. In particular, for β-sheet there is a broad band of dominant frequencies from 15 to 40 cm^{-1}, for α-helix there are a number of discrete peaks from 55 to 115 cm^{-1} and for turns there are peaks from 4 to 73 cm^{-1}. The side-chain χ torsion angles have dominant frequencies above 40 cm^{-1}. Although there is a separation of frequency ranges, the torsion angle vibration of, say, the α-helical residues is still made up of many different modes (at least eight). The lower frequencies of β-sheet compared to α-helix are expected since the β-hairpin is a much softer structure than the α-helix. The higher frequency of the side-chain torsion angles occurs because the side-chains are lower in mass than the elements of secondary structure. The low frequencies (and large amplitudes) of the loops occur because these residues are not subject to strong restraining forces.

The same general features are seen for the other three proteins, with some significant differences for the two larger proteins, ribonuclease and lysozyme for β-sheet there are additional high frequencies above 100 cm^{-1}, for α-helix there are peaks up to 150 cm^{-1}, and for the side-chains the band of frequencies starts above 50 cm^{-1}.

Figure 3 shows the r.m.s. fluctuation of each α-carbon atom, $\sigma_{\alpha i}$, as a function of position along the BPTI polypeptide chain. The results for the TX and TE130 conformations are almost identical indicating their insensitivity to r.m.s. differences of conformation of 2·5 Å (all atoms) and 1·8 Å (α-carbon atoms). The fluctuations are smallest for residues involved in α-helix and β-sheet secondary structure and largest for residues in the exposed turns and at the chain termini. $\sigma_{\alpha i} = \langle \Delta \mathbf{r}_{\alpha i}(\tau)^2\rangle^{1/2}$ is calculated as:

$$\left\{\sum_k^n (\sigma_{\alpha i}^k)^2\right\}^{1/2},$$

where $\sigma_{\alpha i}^k$, the r.m.s. fluctuation of the ith α-carbon atom due to mode k, is given by:

$$\sigma_{\alpha i}^k = (\Delta \mathbf{r}_{\alpha i}^k \cdot \Delta \mathbf{r}_{\alpha i}^k)^{1/2},$$

$$\Delta \mathbf{r}_{\alpha i}^k = \sum_j^n (\partial \mathbf{r}_{\alpha i}/\partial q_j) A_{jk}\alpha_k/2^{1/2}.$$

$\mathbf{r}_{\alpha i}$ is the position vector of the ith α-carbon atom, q_j is the jth torsion angle, A_{jk} are the elements of the matrix that transforms normal mode co-ordinates to torsion angles and α_k is the thermal amplitude of the kth normal mode (see eqn (16)). This calculation assumes a linear dependence between $\mathbf{r}_{\alpha i}$ and q_j that is not valid for large amplitude torsional vibration. At 300 K, relaxing this assumption has a very small effect on $\sigma_{\alpha i}$. Both $\sigma_{\alpha i}$ and $\sigma_{\alpha i}^k$ can be further summed over all the residues i giving averaged r.m.s. fluctuations:

$$\sigma_\alpha = \left\{\sum_i^N (\sigma_{\alpha i})^2/N\right\}^{1/2}$$

and

$$\sigma_\alpha^k = \left\{\sum_i^N (\sigma_{\alpha i}^k)^2/N\right\}^{1/2}.$$

The fluctuations of the ϕ and ψ torsion angles show a different variation with position along the chain (see Fig. 4). Although torsion angles in α-helices show small fluctuations, those in β-sheet are of average magnitude. These results, which are also found for the other three proteins, indicate that a region of chain, like a β-sheet, which has considerable local torsion angle motion need not have large overall α-carbon motion.

(d) *Temperature factors*

The atomic fluctuations calculated for each atom can be compared with fluctuations deduced from the crystal structure analysis. Because the

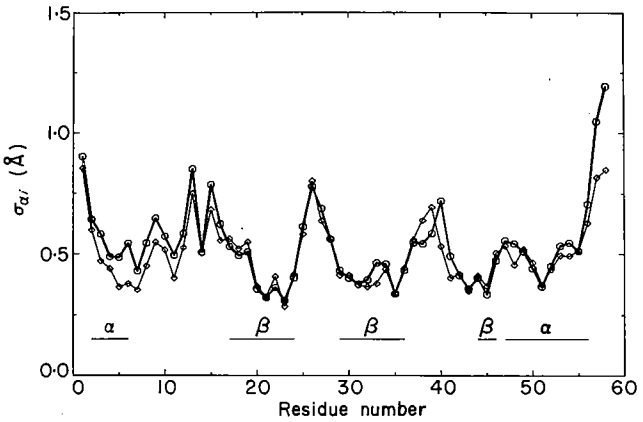


Figure 3. Showing for BPTI how the r.m.s. fluctuation of each α-carbon atom, $\sigma_{\alpha i}$, varies with residue number, i. The vibration is least for residues in α-helix or β-sheet secondary structure and most for residues at the chain termini or the exposed loops. The results shown by the heavy unbroken line with circles for each data point are for the TX conformation; the data points for the TE130 conformation, shown as diamonds connected by a light unbroken line, are almost identical.

crystallographic experiments give temperature factors (or B values) we use these in the comparison (see Table 3); the relationship between temperature factor, B, and atomic fluctuation, σ, is given by $B = 8\pi^2\sigma^2/3$. The magnitude of the α-carbon B values are close to the experimental values in all cases except for lysozyme, where the experimental value is twice as large. This may be due to the lower resolution and different method of B-value refinement used for lysozyme (Sternberg *et al.*, 1979). The agreement for the individual atomic temperature factors is good with correlation coefficients between 0·40 and 0·86.

Figure 5 shows the individual α-carbon temperature factors calculated and observed for all four proteins. The calculated values show the same variation with residue number seen for the $\sigma_{\alpha i}$ (see Fig. 3): regions of α-helix and β-sheet secondary structure have low B values (≈5 Å^2), and turns and chain termini have high values (≈20 Å^2). Many of the regions of high calculated B value also correspond to high observed values. In general, the calculated values are too low in regions of secondary structure and are too high in the turn regions.

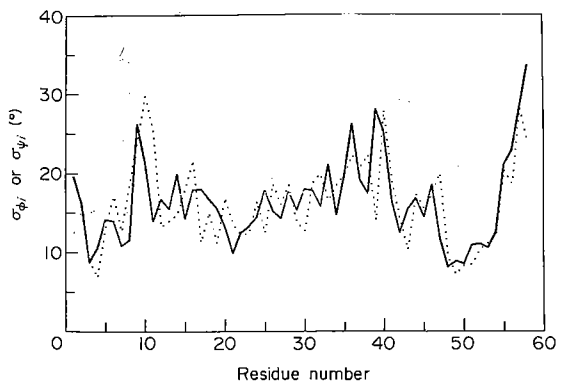


Figure 4. Showing for BPTI how the r.m.s. torsional fluctuations $\sigma_{\phi i}$ (· · · ·) and $\sigma_{\psi i}$ (——) vary with residue number, i. These quantities are calculated by averaging over all normal modes using, e.g. $\sigma_{\phi i} = \{\sum_k^n (\sigma_{\phi i}^k)^2\}^{1/2}$, where $\sigma_{\phi i}^k$, the r.m.s. fluctuation in torsion angle ϕ_i due to mode k is given by $A_{ik}\alpha_k/2^{1/2}$.

(e) *Low frequency motion*

Figure 6 shows the contribution of different modes to the average r.m.s. fluctuation of α-carbon atoms and torsion angles in BPTI. For α-carbons, σ_α^k decreases very rapidly with mode number k: the r.m.s. fluctuation due to the eight lowest modes, $\sigma_\alpha^{1\text{ to }8}$, is >70% of that for all modes, σ_α. For the torsion angles, σ_q^k decreases more slowly with mode number: the r.m.s. fluctuations due to the eight lowest modes is 30% of that for all modes. Similar results are obtained for all the other proteins studied: the number of lowest modes that give 70% of σ_α is between 3 for crambin and 8 for lysozyme (Table 2). In ribonuclease, 70% of σ_α is due to only 0·9% of the modes. It should be noted that because σ is a r.m.s. quantity, it is more appropriate to consider percentage contributions to σ^2. When this is done, a contribution of 70% to σ is seen to be equivalent to a contribution of only 50% to σ^2. High frequency modes that result in very small α-carbon fluctuation still have appreciable torsion angle fluctuation; these torsion angle changes must be locally correlated so as not to move the α-carbon atoms. Because of their dominant contribution to the overall atomic motion of the backbone, the low frequency modes are considered individually in the analysis presented below.

(i) *Backbone motion in BPTI*

Figure 7 shows how the r.m.s. fluctuation of each α-carbon atom, $\sigma_{\alpha i}^k$, varies with position along the

Table 3

Comparison of experimental and calculated temperature factors (B values)†

Protein	Mean (C^α atoms)		Correlation coefficient of obs. and calc.			
	Obs.	Calc.	All§	C^α	Main	Side
BPTI‡ (old)	11·4	8·0	0·54	0·86	0·78	0·40
(new)	14·6	8·0	0·65	0·65	0·65	0·59
Crambin	5·8	7·2	0·60	0·74	0·53	0·55
Ribonuclease	10·6	11·8	0·59	0·57	0·56	0·57
Lysozyme‖	18·0	9·2	—	—	0·56	0·60

† B values are calculated at 300 K.

‡ BPTI has been solved in 2 crystal forms: "old" and "new" (Walter & Huber, 1983). Although the old co-ordinates are used here, the calculated temperature factors are correlated with both sets of experimental B values.

§ All denotes all non-hydrogen atoms; C^α denotes the α-carbon atoms; Main denotes the main-chain atoms N, C^α, C, O; Side denotes the remaining atoms. The correlation coefficient, CC, between sets of data $\{p_i\}$ and $\{q_i\}$ is calculated as

$$CC = (\langle pq\rangle - \langle p\rangle\langle q\rangle)/\{(\langle p^2\rangle - \langle p\rangle^2)(\langle q^2\rangle - \langle q\rangle^2)\}^{1/2}.$$

‖ Experimental B values of lysozyme (M. Sternberg, personal communication) were not refined for individual atoms; instead a single value was used for the main-chain and side-chain atoms of each residue. Thus, it is not possible to give the correlation coefficient for All and C^α. (Since the observed main-chain atoms include the C^β atoms, the calculated values do as well.)

chain for each of the eight lowest frequency modes of BPTI. The variations of $\sigma_{\alpha i}^{1\text{ to }8}$, the total r.m.s. fluctuation of each residue for modes 1 to 8, and of $\sigma_{\alpha i}$, the total r.m.s. fluctuation of each residue for all 208 modes, are also shown. $\sigma_{\alpha i}$ and $\sigma_{\alpha i}^{1\text{ to }8}$ are almost identical in their variation with residue number, i (correlation coefficient 0·97), and both show the same correlation with secondary structure discussed before.

As:

$$\sigma_{\alpha i}^{1\text{ to }8} = \left\{\sum_{k=1}^{8} (\sigma_{\alpha i}^{k})^2\right\}^{1/2},$$

it is of interest to see how the characteristic shape of $\sigma_{\alpha i}^{1\text{ to }8}$ is built up from the $\sigma_{\alpha i}^{k}$ curves. Both $\sigma_{\alpha i}^{1\text{ to }8}$ and $\sigma_{\alpha i}$ have peaks at residues 1, 9, 13, 15, 26, 40 and 58. None of the individual $\sigma_{\alpha i}^{k}$ curves has all these features. The peak at residue 1 comes from modes 3, 4 and 6; that at residue 9 from modes 3 and 4; that at residue 13 from modes 1, 2, 5 and 8; that at residue 15 from mode 1; that at residue 26 from modes 2, 3, 4, 5 and 6; that at residue 40 from modes 1, 4 and 7; and that at residue 58 from modes 3, 5, 6 and 7. Thus, the very reasonable variation of $\sigma_{\alpha i}^{1\text{ to }8}$ with residue number comes from the eight modes acting together; if any one mode were to be omitted, the curve would change noticeably.

In the other three proteins, the r.m.s. α-carbon fluctuation due to the eight lowest modes, $\sigma_{\alpha i}^{1\text{ to }8}$, is also very similar to that due to all modes, $\sigma_{\alpha i}$ (correlation coefficients of 0·99, 0·92 and 0·92 for crambin, ribonuclease and lysozyme, respectively). In each case, the r.m.s. α-carbon fluctuation due to the lowest mode, $\sigma_{\alpha i}^{1}$, is also similar to $\sigma_{\alpha i}$ for all modes (correlation coefficients of 0·82, 0·80 and 0·77 for crambin, ribonuclease and lysozyme, respectively).

The motion of the α-carbon atoms in the eight lowest frequency modes of BPTI is shown in three dimensions in Figure 8. When viewed stereoscopically, the arrows show the nature of the motion clearly. The motion was further investigated by generating 20 conformations for each mode and displaying and filming these conformations on a high-speed vector graphics device.

The general impression obtained from viewing the motion on the computer graphics display is of rocking, twisting, rolling, pivoted rotation and breathing (opposed translation of adjacent segments). These motions involve segments of four or five residues that are about 10 Å in extent. The motions are also collective in that they involve simultaneous motion of all parts of the protein molecule. In the following we give a subjective description of each mode. This gives an intuitive understanding of the nature of the motion that can occur in proteins; the detailed dynamics of each mode will depend on the potential energy functions used. The arrows in Figure 8 give a surprisingly good static representation of the motion, especially when viewed stereoscopically.

Mode 1 involves the rotation of residues 12 to 17 and 37 to 40 at the top of the molecule. Residues 1 to 8 move as a rigid segment from right to left in an arm-swinging motion. Motions of the rest of the molecule are small.

Mode 2 involves a characteristic bending motion in which the top and bottom of the molecule move to the left as the central section moves to the right. This mode would be expected to be one of the dominant low-frequency bending vibrations of an isotropic elastic solid that has the elongated shape of BPTI. There is a noticeable twisting of the β-hairpin.

Mode 3 involves concerted rotation of residues 57 to 58, 1, 25 to 27 and 8 to 9 about a centre close to the centroid of the molecule; the rotation occurs in

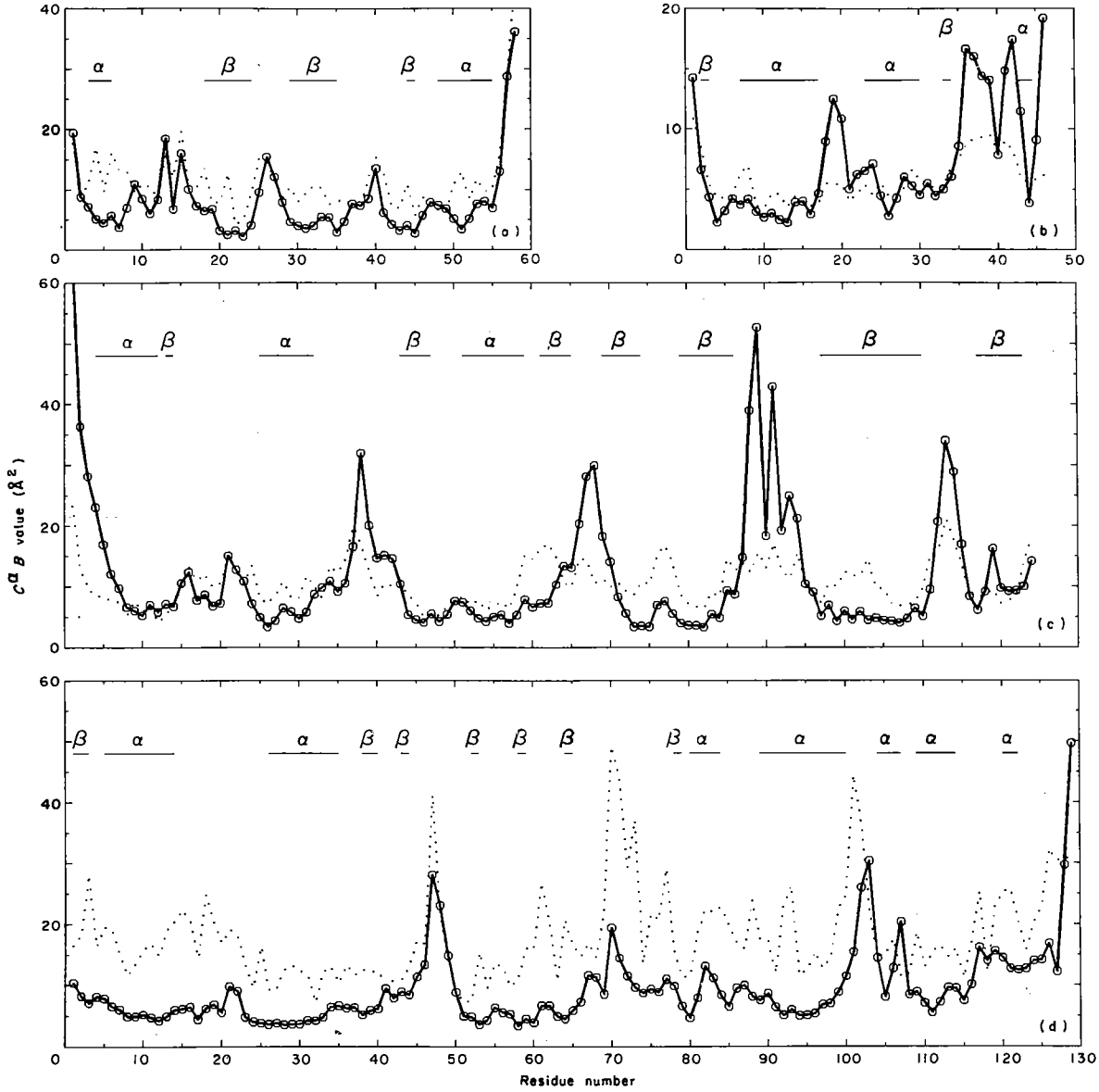


Figure 5. Comparing the experimental (· · · ·) and calculated (———) variations with residue number of the α-carbon B values (temperature factors). The ith calculated B value is derived from the r.m.s. fluctuation of the ith α-carbon atom when all modes are excited at 300 K. The 4 graphs show: (a) BPTI, (b) crambin, (c) ribonuclease and (d) lysozyme. Regions of α-helix and β-sheet secondary structure (Kabsch & Sander, 1983) are shown as unbroken lines.

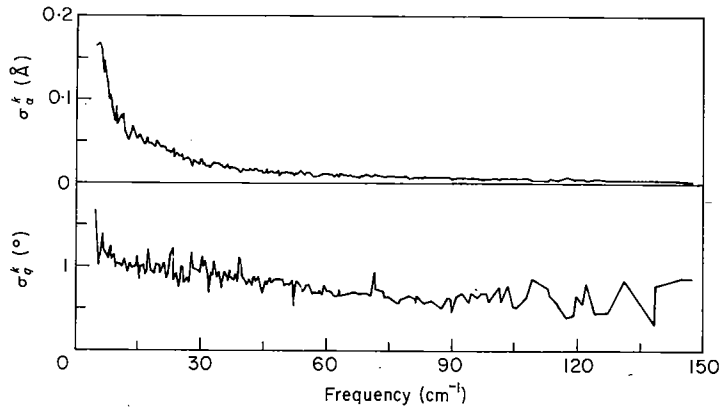


Figure 6. Showing for BPTI how the average r.m.s. fluctuations of the α-carbon atoms, σ_α^k, and single-bond torsion angles, σ_q^k, depend on the frequency of the mode k. The value of σ_α^k drops rapidly as the frequency increases, with $\sigma_\alpha^{208}/\sigma_\alpha^1 = 0{\cdot}005$. This means that the low frequencies make the major contribution to σ_α, the average r.m.s. C^α fluctuation due to all modes. On the other hand, σ_q^k drops rather slowly as the frequency increases, with $\sigma_q^{208}/\sigma_q^1 = 0{\cdot}08$. This means that the average r.m.s. torsion angle fluctuation due to all modes, σ_q, depends on vibrations over a wide range of frequencies.

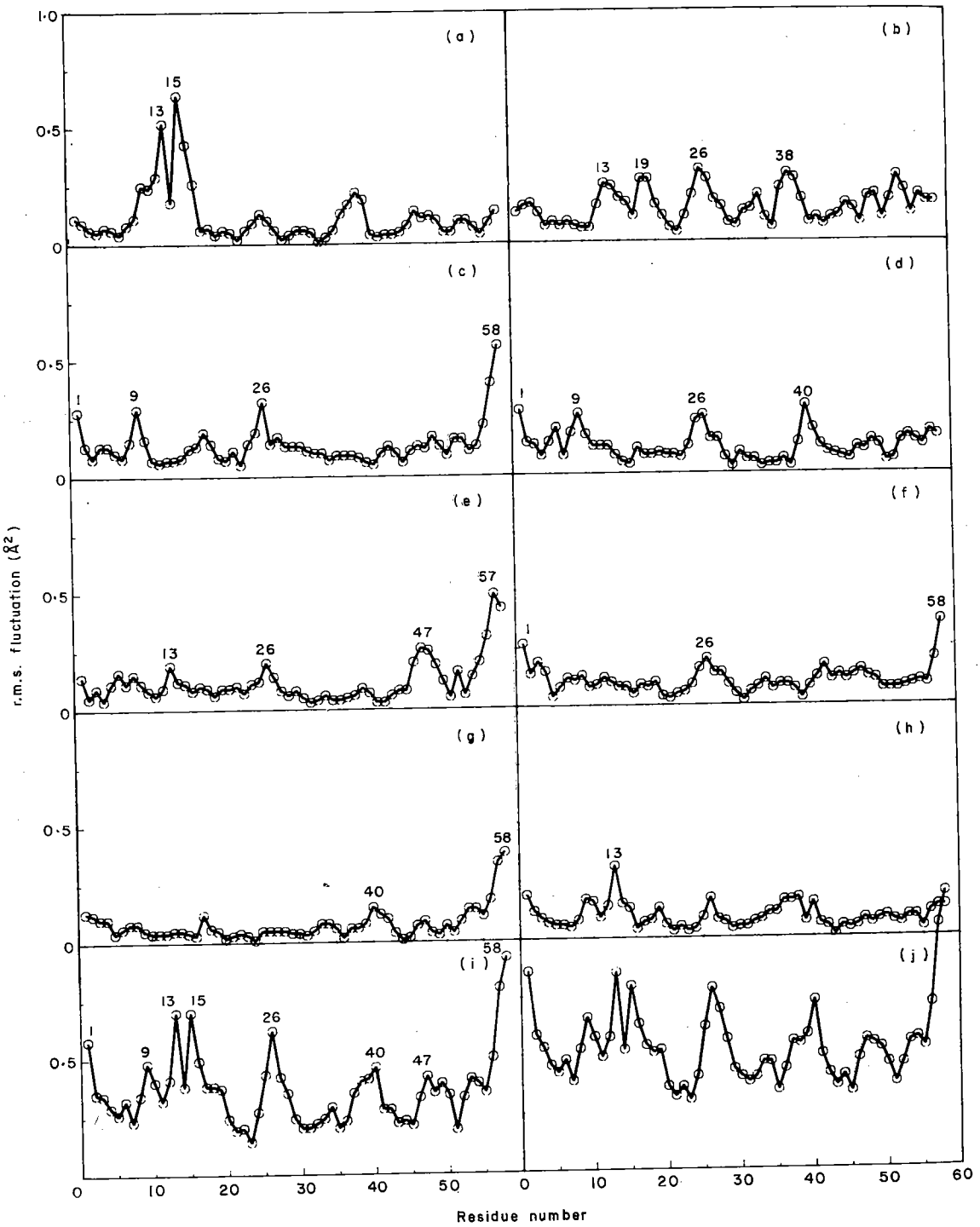


Figure 7. Showing for BPTI the variation with residue number i of the r.m.s. fluctuation of each α-carbon atom, $\sigma_{\alpha i}^{k}$, as a result of the 8 lowest frequency modes ($k = 1$ to 8 in (a) to (h), respectively). The total r.m.s. fluctuation for these 8 modes together, $\sigma_{\alpha i}^{1\text{to}8} = \{\sum_{k=1}^{8} (\sigma_{\alpha i}^{k})^2\}^{1/2}$, and the total r.m.s. fluctuation for all 208 modes, $\sigma_{\alpha i} = \{\sum_{k=1}^{208} (\sigma_{\alpha i}^{k})^2\}^{1/2}$, are also shown (in (i) and (j), respectively). Note that while $\sigma_{\alpha i}^{1\text{to}8}$ is very similar to $\sigma_{\alpha i}$, none of the individual $\sigma_{\alpha i}^{k}$ curves has all the features of $\sigma_{\alpha i}$. The frequencies of these 8 modes are given in the legend to Fig. 8.

the plane of the drawing and there is little other motion.

Mode 4 involves two coupled breathing motions. Residues 9 to 10 and 40 to 41 move in opposite directions as do residues 25 to 26 and 9 to 10. When the cleft between 9 to 10 and 40 to 41 closes, that between 25 to 26 and 9 to 10 opens.

Mode 5 involves rotation of the entire C-terminal helix about a horizontal axis that passes through the helix centroid; as residues 46 to 50 move down, residues 54 to 58 move up.

Mode 6 involves a scissors-like motion of residues 24 to 29 and 57 to 58 about a pivot axis perpendicular to the plane of the drawing and

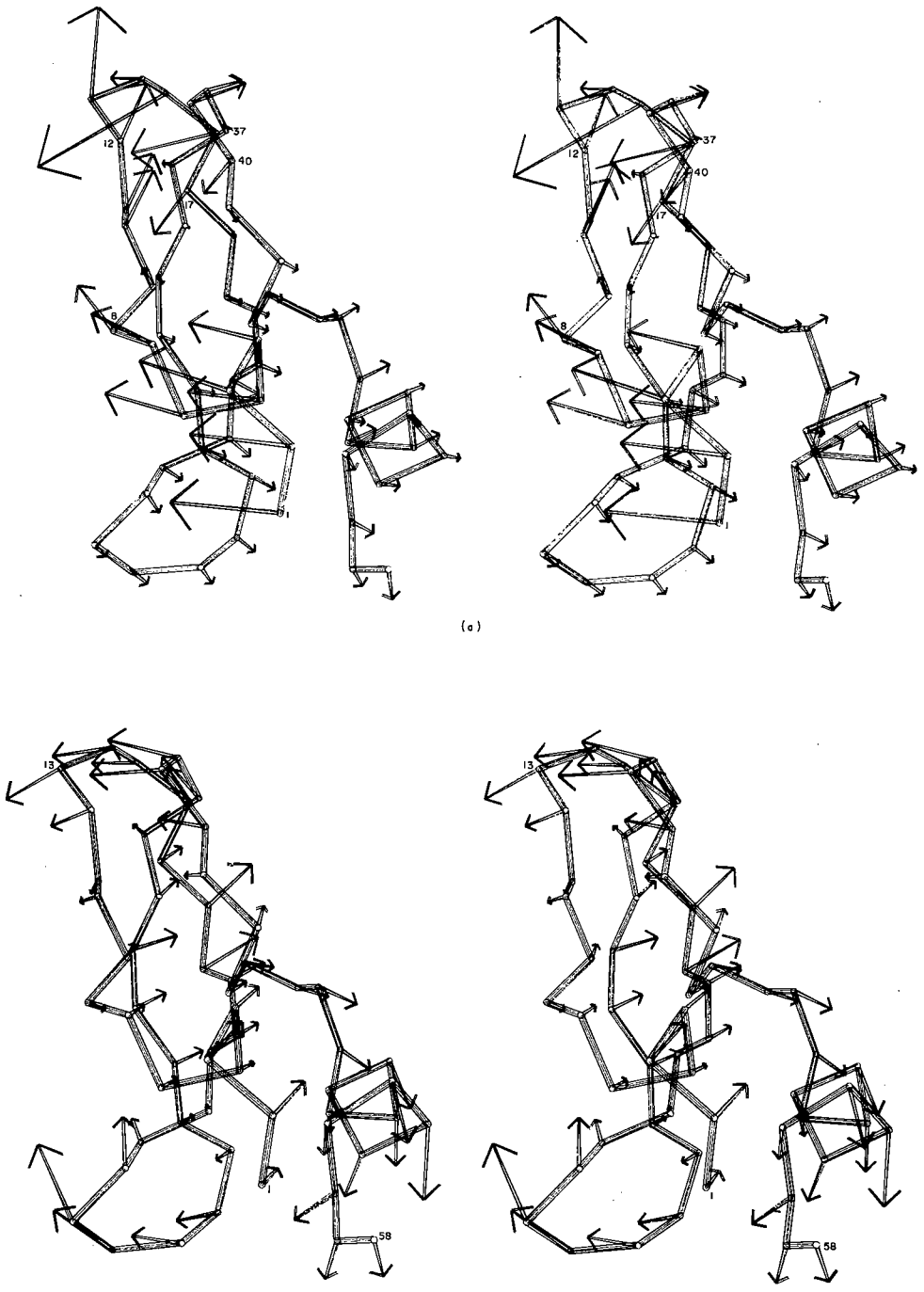


Fig. 8

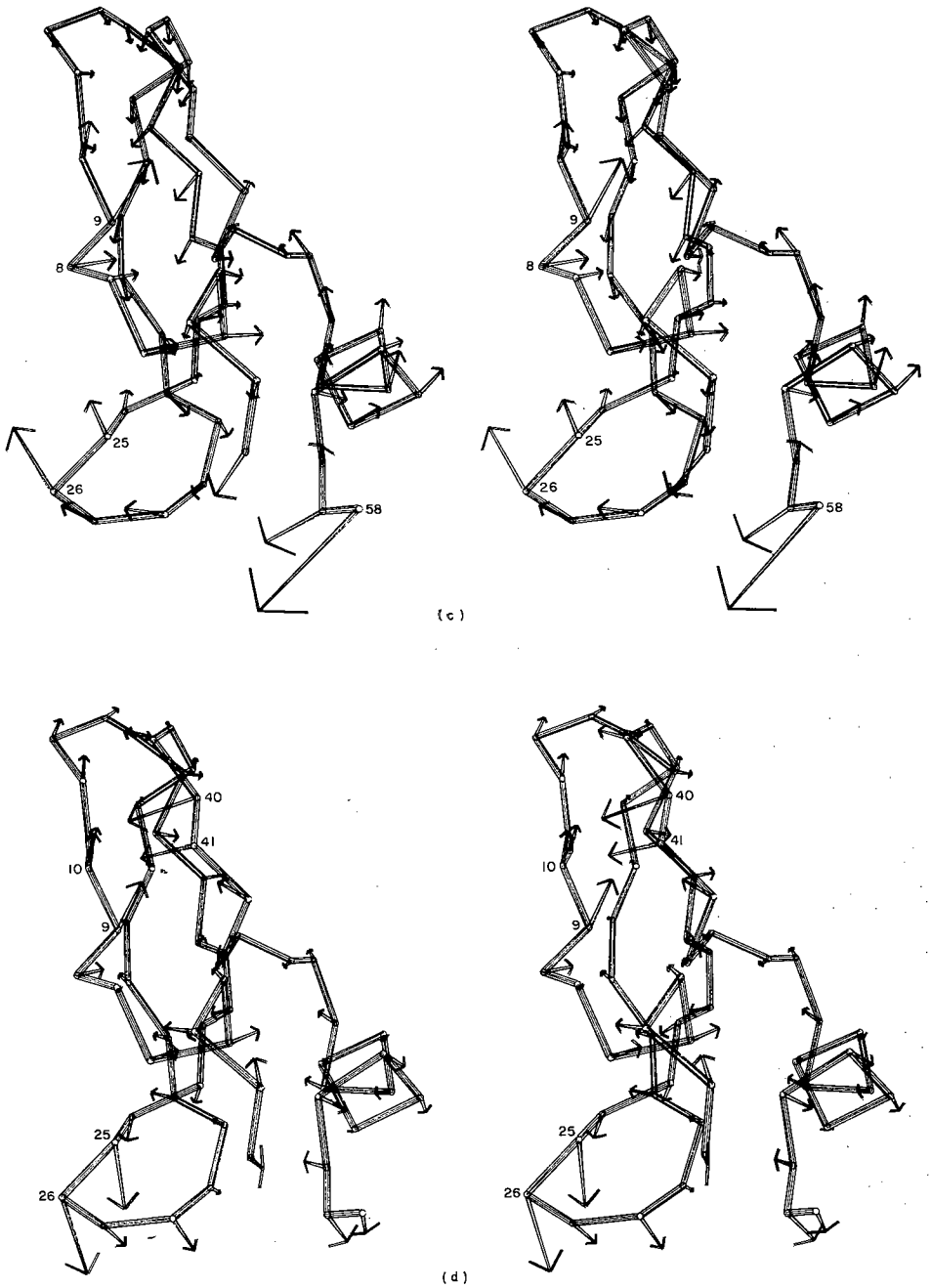
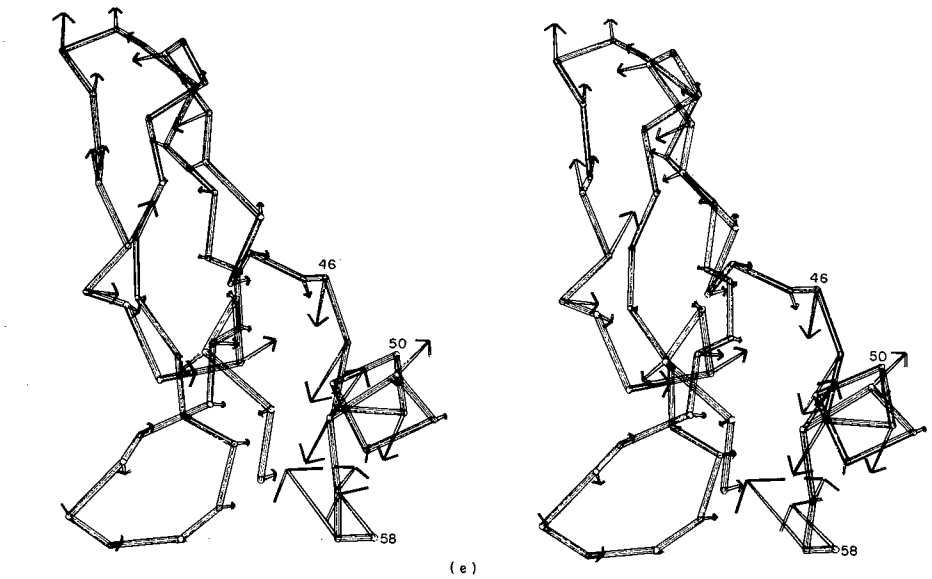
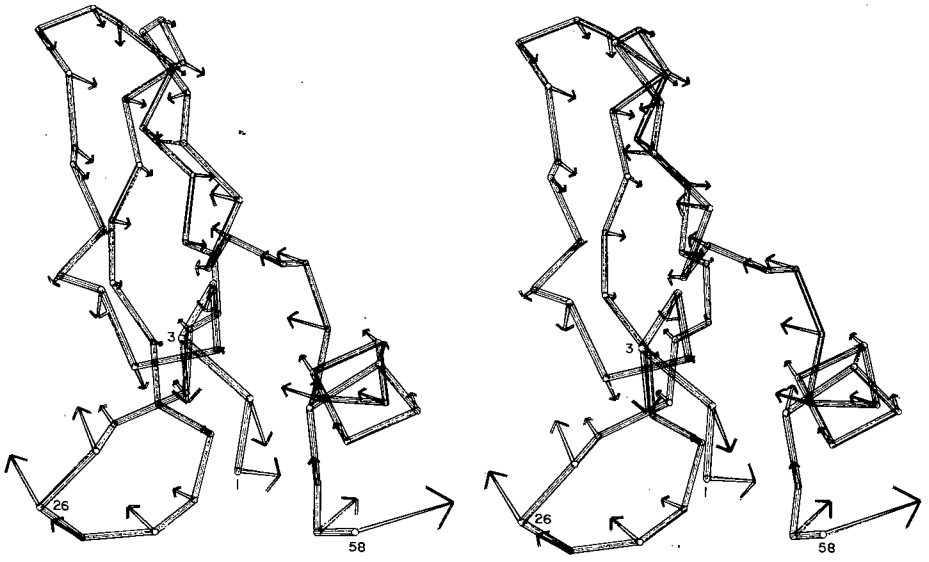


Fig. 8, *cont.*



(e)



(f)

Fig. 8, *cont.*

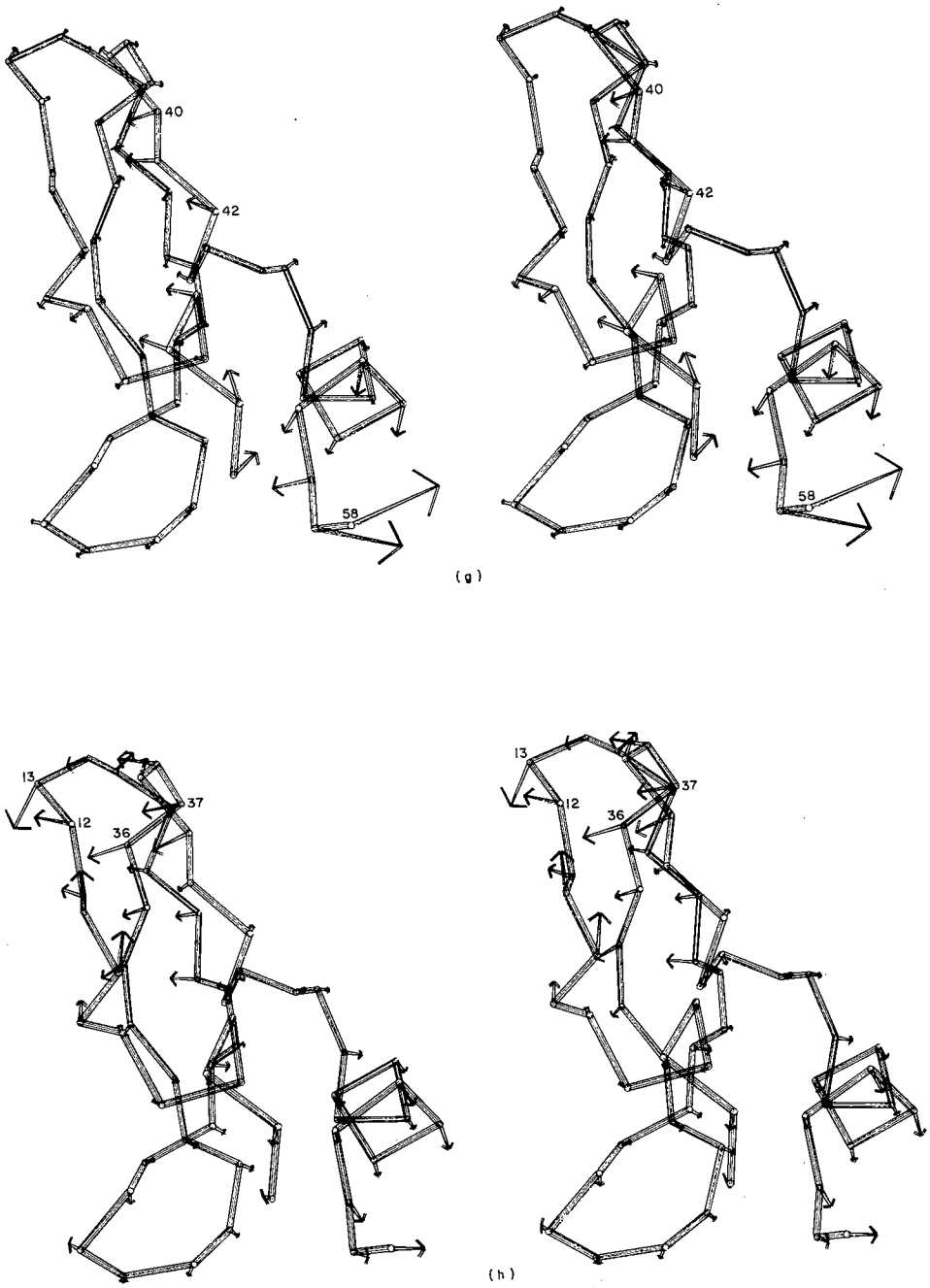


Figure 8. Showing the α-carbon atom shifts in BPTI for each of the 8 lowest frequency modes. The α-carbon positions are those in the TX minimum energy conformation. The arrows show the direction and extent of motion about the equilibrium positions; the end point of the arrow on the ith α-carbon atom, $\mathbf{r}_{\alpha i}$, is given by $\mathbf{r}'_{\alpha i} = \mathbf{r}_{\alpha i} + 7\Delta\mathbf{r}^k_{\alpha i}$, where the factor of 7 is introduced to ensure that all the arrows are visible. $\Delta\mathbf{r}^k_{\alpha i}$ is the shift in the position of the ith α-carbon atom caused by changing all the torsion angles q_j by $\Delta q_j = A_{jk}\alpha_k$, their maximum displacement in the kth normal mode. These shifts do not assume a linear dependence of $\mathbf{r}_{\alpha i}$ on q_j as is assumed in calculating $\Delta\mathbf{r}^k_{\alpha i} = \sum_j^n (\partial\mathbf{r}_{\alpha i}/\partial q_j) A_{jk}\alpha_k$ (e.g. in Fig. 7). For BPTI, this assumption only affects the motions of residues 6 to 11 in mode 1 (cf. Fig. 8(a) with Fig. 7(a)). The motion of each atom is generally symmetrical about the equilibrium position, but the shifts in the reverse direction are not shown for greater clarity. Because the motion does not translate or rotate the centre of mass, the arrows add vectorially to zero. The modes have periods ranging from 7·3 to 4·2 ps with frequencies: (a) $\nu_1 = 4{\cdot}56\ \mathrm{cm}^{-1}$, (b) $\nu_2 = 5{\cdot}40\ \mathrm{cm}^{-1}$, (c) $\nu_3 = 5{\cdot}92\ \mathrm{cm}^{-1}$, (d) $\nu_4 = 6{\cdot}41\ \mathrm{cm}^{-1}$, (e) $\nu_5 = 6{\cdot}56\ \mathrm{cm}^{-1}$, (f) $\nu_6 = 7{\cdot}33\ \mathrm{cm}^{-1}$, (g) $\nu_7 = 7{\cdot}75\ \mathrm{cm}^{-1}$, (h) $\nu_8 = 7{\cdot}93\ \mathrm{cm}^{-1}$. In each case the residues that move are numbered. The nature of the motion in each mode is described more fully in the text.

through the molecular centroid. Residues 1 to 3 move down, allowing better packing of the loop and C terminus.

Mode 7 involves the wagging of the C terminus, with large movements of residues 56 to 58, the tail, and smaller movements of the entire α-helix. The motions of residues 40 to 42 oppose the motions of residues 56 to 58.

Mode 8, the last mode described in detail, involves large motions of residues 12 to 14 and 36 to 38. The overall impression is one of overall twisting with the top and bottom of the molecule moving together.

It is of interest to assess the extent to which the nature of the motion in the individual modes depends on the conformation used for the normal-mode dynamics. Conformations TX and TE130 of BPTI differ significantly (2·5 Å r.m.s. for all atoms and 1·8 Å r.m.s. for α-carbon atoms). The α-carbon fluctuations due to the eight lowest modes, $\sigma_{\alpha i}^{1\text{ to }8}$, or to all the modes, $\sigma_{\alpha i}$, are very similar for both conformations. The fluctuations due to the single modes appear less similar.

Correlation coefficients are calculated for the α-carbon atom shifts, $\Delta\mathbf{r}_{\alpha i}^{k}$, of each of the eight lowest modes of TX and TE130. Each mode of TE130 is correlated with only one mode of TX (correlation coefficient of between 0·3 and 0·7). For example, the lowest mode of TE130, with a frequency of 3·8 cm^{-1}, is very similar to the second lowest mode of TX, with a frequency of 5·4 cm^{-1}. This mode is very characteristic in that it involves overall bending and twisting of the β-sheet. The eight low frequency modes of the two conformations of ribonuclease, TX and TEX (1·2 Å r.m.s. deviation for all atoms), are also correlated (correlation coefficients between 0·30 and 0·82), but there is not always a one to one correspondence between modes and again the same type of motion does not necessarily have the same frequency.

Correlation coefficients are also calculated between the α-carbon shifts, $\Delta\mathbf{r}_{\alpha i}^{k}$, of the eight lowest modes in the same structure. Generally, there is little correlation, indicating that these modes are mostly independent modes of motion in Cartesian co-ordinates. This orthogonality of the shift vectors is not a necessary consequence of a calculation in torsion angle co-ordinates.

(ii) *Domain motion in ribonuclease and lysozyme*

Figure 9 illustrates the α-carbon atom motion in the lowest frequency modes of ribonuclease and lysozyme. In both cases the motion is very characteristic, involving an opening and closing of the active-site cleft.

For ribonuclease, the two parts of the large β-sheet (upper and lower in Fig. 9(a)) twist in opposite directions. This twisting motion is particularly clear for the lower part. The largest shifts are for residues 1, 38, 68 and 88, which are at the extremes of the cleft. In this mode of motion, the three helices in ribonuclease (residues 4 to 12, 24 to 33 and 50 to 58) move as rigid bodies with all the shift arrows on a particular helix pointing in the same direction. The motion can be represented by rigid body motion of large fragments: the lower part (in Fig. 9(a)) consisting of residues 20 to 46 and 80 to 102 and the upper part consisting of residues 47 to 79 and 103 to 124. These two fragments are not like the folding units or domains into which large proteins are divided (Wetlaufer, 1973), as more than one length of polypeptide chain is involved in each fragment.

For lysozyme, the motion of the upper and lower parts of the molecule (Fig. 9(b)) is more complicated. The upper part (residues 40 to 86), which consists of a β-sheet and exposed loops, deforms as it bends down to close the cleft. This is evidenced by a change in direction of the shift arrows across the sheet. The lower part (residues 1 to 39 and 87 to 129), which consists of five α-helices, appears to be rotating, with each helix moving as a rigid body. The largest shifts are for residues 47, 70 and 103, which are at the extremes of the cleft. One of the two fragments that move as more or less rigid bodies in lysozyme would be classified as a folding domain (upper fragment, residues 40 to 86), while the other would not be.

(f) *Correlated motions*

One of the major advantages of normal-mode dynamics over molecular dynamics is the ability to calculate correlation coefficients. These quantities can be calculated accurately only by averaging over very long dynamics trajectories, whereas they are obtained with analytical precision from normal-mode analysis.

The motion of all atoms in a single mode are by definition in phase and completely correlated. When many modes are added together, there is interference that reduces the extent of the correlation. Nevertheless, even when all modes are vibrating with different phases at room temperature, there are significant correlations.

(i) *Correlated vibration of α-carbon atoms*

Figure 10 shows the correlation coefficients of all pairs of α-carbon shifts $\Delta\mathbf{r}_{\alpha i}$ and $\Delta\mathbf{r}_{\alpha j}$ in BPTI. The highest correlation occurs between adjacent residues, where the correlation coefficient is between 0·5 and 0·8. The correlation of the motion of next-nearest neighbours along the chain is lower but significant. Residues that are connected by hydrogen bonds in the 3_{10}-helix (residues 2 to 6), the β-sheet (residues 16 to 24, 29 to 36, 44 to 46) and the α-helix (residues 47 to 56) also show positive correlation. Residues that are close in space but not connected by hydrogen bonds are often negatively correlated, indicating that they move in opposition. This is particularly marked for the residues near the disulphide bonds that connect the α-helix to the 3_{10}-helix (5 to 55) and to the β-sheet (30 to 51). Although there are few positive correlations between residues more than 10 Å apart,

Figure 9. Showing the domain motion due to the lowest frequency modes in ribonuclease and lysozyme. The arrows on the α-carbon backbone are drawn as described in the legend to Fig. 8. (a) Ribonuclease, $\nu_1 = 2{\cdot}43\ \mathrm{cm}^{-1}$, period = 13·7 ps. (b) Lysozyme, $\nu_1 = 2{\cdot}98\ \mathrm{cm}^{-1}$, period = 11·2 ps.

there are significant negative correlations between residues as far apart as 20 Å.

(ii) *Correlation of torsion angles*

Figure 11 shows the correlation coefficients for all 208 torsion angles of BPTI. Because the angles are ordered ϕ_i, ψ_i, ϕ_{i+1}, ψ_{i+1}, ..., χ_i, ..., the upper block gives the main-chain/main-chain correlations, the lower block gives the side-chain/side-chain correlations, and the off-diagonal block gives the main-chain/side-chain correlations. The pattern of main-chain correlations is very like the contact map (Fig. 10, lower triangle), showing that the main-chain torsion angles are correlated only for residues close in space. This same pattern is seen for the main-chain/side-chain correlations but not for the side-chain/side-chain correlations. This indicates that the torsion angles of side-chains packed closely together do not vibrate in unison.

The correlated vibration of torsion angles in the β-hairpin and α-helix are considered in more detail. For the β-hairpin (Table 4) there are two types of strong correlations: (1) intrastrand, in which the correlated angles are from adjacent residues in the

Table 4

Torsion angle shift correlations in the BPTI β-hairpin

	Arg		Ile		Ile		Arg		Tyr		Phe		Tyr		Asn		Ala		Lys		Ala		Gly		Leu		Cys		Gln		Thr		Phe		Val		Tyr		Gly	
	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	7	7	8	8	9	9	0	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	0	0	1	1	2	2	3	3	4	4	5	5	6	6
17 Arg		2	−2	−2	1																														1	−2	−2	1	3	
17 Arg	2		−8			−2	1																												−1	1	−2	−1	1	1
18 Ile	−2	−8			1	−1			−1																										−1	2	−1	3	−2	
18 Ile	−2			−7		−1		−1			1																			1		−1	1	−1	2	−3	−1		1	
19 Ile	1		1	−7									−1																	−1	−1		−1	1	−3	3	−1	2	−1	
19 Ile		−2	−1					−9	−2	2	−1		−1																	−1			2	−1			1	−1			
20 Arg		1		−1		−9			3	−2	1	−1	1																			−2	2		−1	−1	1			
20 Arg						−2	3		−8	1			1																−1	−2	4		−1	−2	1	1	−2				
21 Tyr			−1	−1		2	−2	−8		−1	−1	−1																		3	−4	1	1	3	−2	−1	1			
21 Tyr						−1	1	1	−1		−8	3	−2														1					1	−1	−1	1					
22 Phe							−1		−1	−8		−2	−1			1				1				1		−2	1	1	−1	2	−2	1						1		
22 Phe				1		−1	1	1	−1	3	−2		−8	2	−1	−1			1	−1			1	−1		1	1	−1	1		1			1						
23 Tyr					−1					−2	−1	−8		−2				−1	−1			−1	1			1	−1	1	−1	1									
23 Tyr											2	−2		−8											1	−2	1	−1		−1	1								
24 Asn												−1		−8		1	−1	1	−2	1	1	−1	1	1		−3	3	−1	−1											
24 Asn											1	−1			1		−7	−2	−1	2	−1		−3	2			−2		2	−1		1	−1						
25 Ala														−1	−1	−7		−3	2	−3	6	−1	1	−3		2	−2												
25 Ala															1	−2	−3		−7	1	−4	1	1																
26 Lys											1			−2	−1	2	−7		−4				1			1	1	−1											
26 Lys											1	−1	−1		1	2	−3	1	−4		−5		−1	−2	−1	1	−2												
27 Ala													−1		1	−1	6	−4		−5		−3	1	−2		2	−2												
27 Ala														−1		−1	1			−3		−9	2		−1	1	1	−1											
28 Gly															1	−3	1	1		−1	1	−9		−3	2	1		1	−1	1									
28 Gly												1	−1		1	2	−3		1	−2	−2	2	−3		−7	−3	2	−1			−1								
29 Leu											1	−1	1						−1			2	−7			−1	−1	2											
29 Leu														1	−3		2			1	2	−1	1	−3			−8	−2	2	−1									
30 Cys										1	−2	1	1	−2	3		−2		1	−2	−2	1		2	−1	−8		3	−4		1								
30 Cys											1	1	−1	1	−1	−2			1			1	1	−1	−1	−2	3		−9	1	−1								
31 Gln								−1			1	−1	1	−1	−1	2		−1			−1	−1	−1		2	2	−4	−9												
31 Gln								−2	3		−1	1	−1			−1							1		−1		1		−9	2									
32 Thr				1	−1	−1		4	−4		2		1	−1		1								−1			1	−1		−9		−2	−1	−2	1	1	−1			
32 Thr					−1				1	1	−2	1		1		−1														2	−2		−9	3	−1	−1				
33 Phe			−1					−1		1	−1	1																			−1	−9				1		−1		
33 Phe				1	−1	2	−2	−2	3	−1																					−2	3		−9		−1	−1				
34 Val		−1		−1	1	−1	2	1	−2	1			1																		1	−1		−9					−1	
34 Val	1		1	−1	2	−3			1	−1																					1	−1				−9	−1	−1	2	
35 Tyr	−2	−2		2	−3	3		−1	−2	1																					−1		1	−1		−9		3	−1	−3	
35 Tyr	−2	−1	−1	−1	−1		1	−1																									−1		−1	3		−9			
36 Gly	1	1	3			2	−1	1			1																							−1			−1	−1	−9		−3	
36 Gly	3		1	−2	1	−1																													−1	2	−3		−3		
	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	7	7	8	8	9	9	0	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	0	0	1	1	2	2	3	3	4	4	5	5	6	6

The correlation coefficients have been scaled by multiplying them by 10.

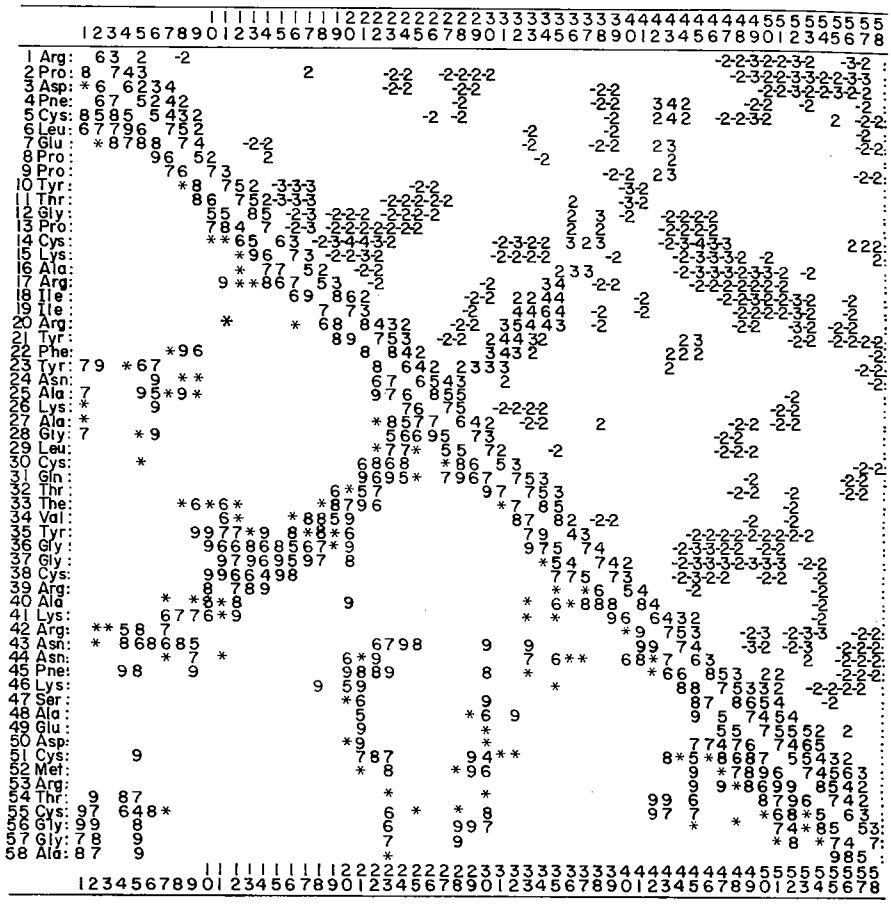


Figure 10. Showing the correlation coefficients of all pairs of BPTI α-carbon atoms (upper diagonal) together with the α-carbon contact map (lower diagonal). The correlation coefficient is calculated as $C_{ij} = \langle \Delta \mathbf{r}_{\alpha i}(\tau) \Delta \mathbf{r}_{\alpha j}(\tau) \rangle / \sigma_{\alpha i} \sigma_{\alpha j}$. C_{ij} is 1 if α-carbon atoms i and j move in the same direction for all modes (perfectly correlated), -1 if they move in opposite directions for all modes (perfectly anticorrelated), and 0 if they move at right angles or there is no relationship between the motions in different modes (uncorrelated). Only those C_{ij} values greater in absolute value than 0·15 are shown and values are multiplied by 10 to avoid the decimal point. The contact map shows the distance, r_{ij}, between α-carbon atoms i and j if this distance is less than 11 Å; if the distance is between 10 and 11 Å it is shown as *. The diagonal along which $C_{ii} = 1$ and $r_{ii} = 0$ is left blank for greater clarity. Notice how most of the α-carbon atoms that are closer in space than 8 Å show correlated motion ($C_{ij} > 0{\cdot}15$). This means that these segments move together as more-or-less rigid bodies for most of the dominant modes. The motion between segments is generally negatively correlated, showing that pairs of segments tend to move in opposite directions for most of the dominant modes.

Table 5

Characteristic correlations of BPTI α-helix and β-sheet

Relative residue no. j	Correlation coefficient for α-helix $\langle \Delta\phi \Delta\phi_j \rangle$	α-helix $\langle \Delta\psi \Delta\psi_j \rangle$	α-helix $\langle \Delta\phi \Delta\psi_j \rangle$	β-sheet $\langle \Delta\phi \Delta\phi_j \rangle$	β-sheet $\langle \Delta\psi \Delta\psi_j \rangle$	β-sheet $\langle \Delta\phi \Delta\psi_j \rangle$
−4	0·04	0·08	−0·08	0	0	0
−3	0·05	0·09	−0·11	0	0	0
−2	0·35	0·25	−0·24	0	0	0
−1	0·21	0	−0·73	0	0	−0·82
0	1·00	1·00	−0·30	1·00	1·00	0
1	0·21	0	−0·29	0	0	0
2	0·35	0·25	0	0	0	0
3	0·05	0·09	−0·09	0	0	0
4	0·04	0·08	0	0	0	0

The correlation coefficients are the average values found for the α-helix and β-sheet of BPTI.

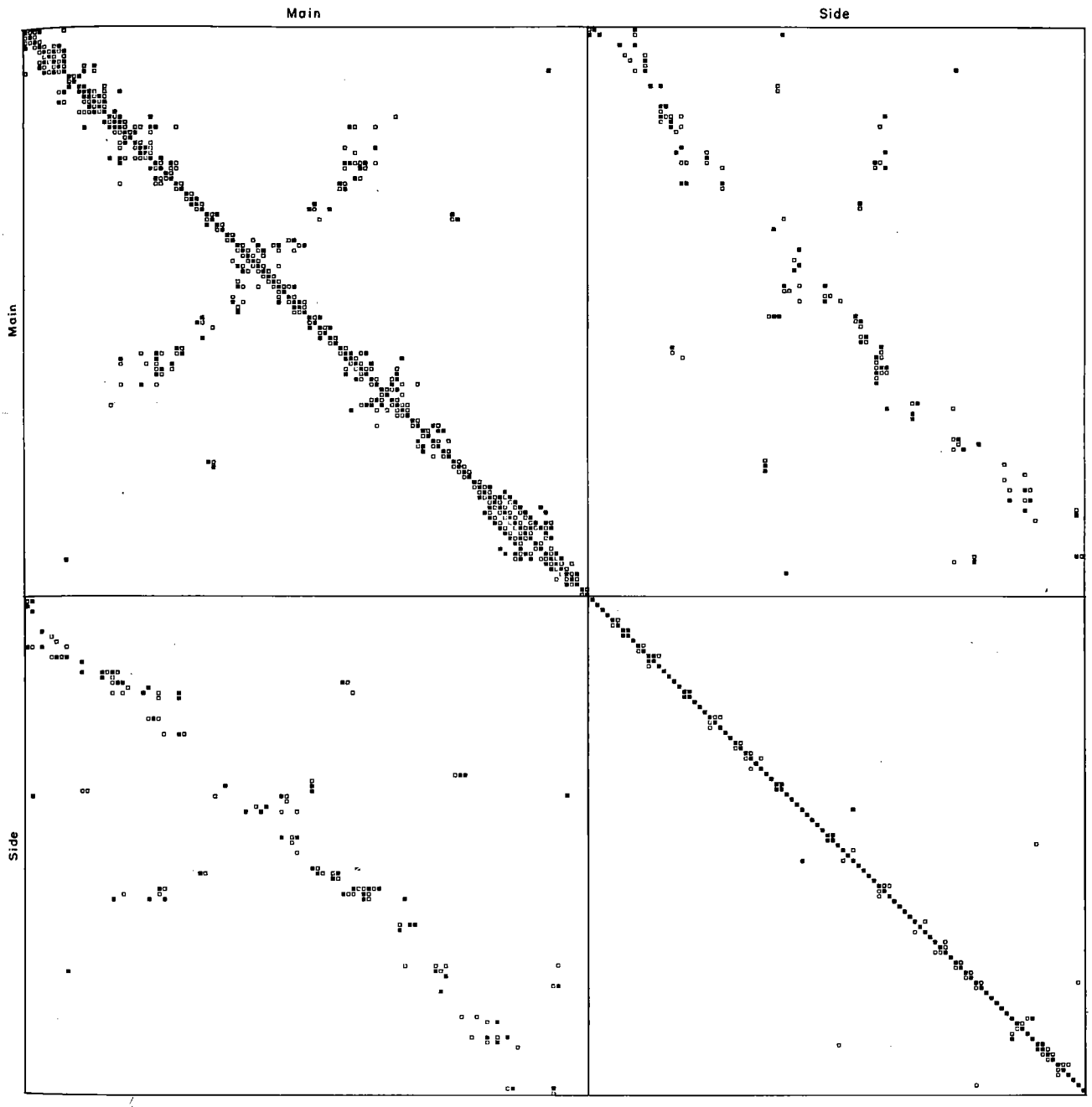


Figure 11. Showing the correlation coefficients of all pairs of BPTI torsion angles $C_{ij}^{q} = \langle \Delta q_i(\tau) \Delta q_j(\tau) \rangle / \langle \Delta q_i(\tau) \rangle^{1/2} \langle \Delta q_j(\tau) \rangle^{1/2}$ (see eqn (14)). The torsion angles, q, are ordered as follows: ψ_1, ϕ_2, ψ_2, ..., ψ_{57}, ϕ_{58}, ψ_{58}, χ_1, χ_2, ... Positive correlations with $C_{ij} > 0{\cdot}15$ are shown as filled squares, whereas negative correlations with $C_{ij} < -0{\cdot}15$ are shown as open squares. The matrix of correlations is divided into blocks corresponding to the main-chain and side-chain angles. The upper block on the diagonal is for main-chain/main-chain correlations, the lower block on the diagonal shows side-chain/side-chain correlations, and the off-diagonal blocks show main-chain/side-chain correlations.

same strand; and (2) interstrand, in which the correlated angles are in different strands. The only significant intrastrand correlation is $\Delta\phi\Delta\psi_{-1} = -0{\cdot}8$. The significant interstrand correlations do not seem regular. For example, there is positive correlation greater than 0·3 for the following pairs of angles: $\phi_{25}\phi_{27}$, $\phi_{24}\phi_{30}$, $\phi_{21}\psi_{31}$, $\phi_{21}\psi_{33}$, $\psi_{20}\phi_{32}$, $\phi_{19}\phi_{35}$, $\phi_{18}\phi_{36}$ and $\phi_{17}\psi_{36}$. Although most pairs involve angles of the same type, there is no simple relationship to the β-hairpin hydrogen bonding that occurs between residue pairs (24,29), (22,31), (20,33) and (18,35).

For the α-helix (Table 5) there are a number of significant correlations. The $\Delta\phi\Delta\phi_j$ and $\Delta\psi\Delta\psi_j$ correlations are positive for $j = -4$ to 4 whereas the $\Delta\phi\Delta\psi_j$ correlation is negative. All the ϕ angles in an α-helix change together as do all the ψ angles; the ϕ angles decrease as the ψ angles increase. Close study of Table 5 shows other interesting features of the torsion angle correlations in α-helices. The largest

values of $\Delta\phi\Delta\phi_j$ and $\Delta\psi\Delta\psi_j$ both occur for $j=\pm 2$ (except from the obvious correlation of 1 at $j=0$). This occurs because the N–C^α bonds of residues i and $i+2$ are almost perfectly antiparallel. The largest negative correlation occurs for $\Delta\phi\Delta\psi_{-1}$. These two angles are attached to the same rigid peptide group. If ϕ_i and ψ_{i-1} are changed in opposite directions, the peptide tilts without affecting the chain path. In the β-sheet, the only significant correlation is $\Delta\phi\Delta\psi_{-1}$, also seen in the α-helix. It is striking how much of an effect changing the conformation from α-helix to β-sheet has on the pattern of correlations. Similar correlations were found in studies of an isolated α-helix (Gō & Gō, 1976; Levy & Karplus, 1979).

4. Discussion

(a) *Applicability of normal-mode dynamics*

(i) *Relation to other methods*

Normal-mode analysis has been used to study molecular vibrations for many years. In most studies of small molecules, the second derivative matrix in internal co-ordinates, **F**, is defined directly in terms of the force constants for each type of interaction. For example, if there is a bond length, b, and bond angle, θ, that have an energy term:

$$E = 1/2K_b(b-b_0)^2 + 1/2K_\theta(\theta-\theta_0)^2,$$

then the energy is a minimum at $b=b_0$ and $\theta=\theta_0$ and the second derivatives are $\partial^2E/\partial b^2 = K_b$, $\partial^2E/\partial\theta^2 = K_\theta$, $\partial^2E/\partial b\,\partial\theta = 0$. This method will not work when there are energy terms that are not simple quadratic functions of the internal co-ordinates b and θ. In large molecules, there are many non-bond interactions that are non-quadratic functions of interatomic distances, which in turn are complicated non-quadratic functions of bond lengths, bond angles and torsion angles. In this case, the energy function will only approximate a quadratic function when the first derivatives are zero, i.e. the energy function has a minimum value. If it is necessary to perform energy minimization, there is no benefit to be gained from using internal co-ordinates (b, θ, ϕ) rather than Cartesian co-ordinates (x, y, z): in both cases the second derivative matrix is a complicated function of the co-ordinates. In fact, the analytical expressions for derivatives are simpler in Cartesian co-ordinates.

In their pioneering studies, Lifson & Warshel (1968) calculated vibrational frequencies by normal-mode analysis using a Cartesian co-ordinate second derivative matrix calculated at the minimum energy conformation. This same method has been used to calculate the normal modes of an α-helix (Levy & Karplus, 1979) and the protein BPTI (Brooks & Karplus, 1983). Because BPTI has so many Cartesian co-ordinates (1740), a minimum energy conformation is not reached and the second derivative matrix is not positive definite, having seven negative eigenvalues.

Torsion angle co-ordinates were used in another normal-mode calculation on the protein BPTI (Gō *et al.*, 1983). In this independent study, the second derivative matrix, **F**, and kinetic energy matrix are calculated by elegant analytical methods (Noguti & Gō, 1983*a*,*b*), after the energy has been minimized exactly. Although it is now well established that normal modes must be calculated at a minimum energy conformation, many studies still construct the second derivative matrix from force constants. Such studies are particularly suspect for large systems that are dominated by hydrogen bonds, van der Waals' interactions and electrostatic forces.

(ii) *Relation to other results*

The results of the two other studies of BPTI normal modes can be compared to the present work. Gō *et al.* (1983) use torsion angle co-ordinates as is done here, but they use a different potential energy function, include all polar hydrogen atoms and have more degrees of freedom (241 as against 208). Although their variation of the r.m.s. main-chain fluctuation with residue number has the same shape as ours (correlation coefficient of 0·88), their average fluctuation is much smaller (0·31 Å as against 0·58 Å). Their lowest frequency of $5{\cdot}7\ cm^{-1}$ is a little higher than ours ($4{\cdot}6\ cm^{-1}$), but their frequency distribution is shifted to higher frequency by about $10\ cm^{-1}$. The energy functions used by Gō *et al.* seem to be at least twice as stiff as those used here, leading to smaller fluctuations and higher frequencies.

Brooks & Karplus (1983) use Cartesian co-ordinates in a very difficult computation that involves diagonalization of a matrix with 1740×1740 elements. Their potential energy function differs from ours and they include all polar main-chain hydrogen atoms and the four internal water molecules. Although their variation of the r.m.s. main-chain fluctuation with residue number is similar to ours (correlation coefficient of 0·84), their average fluctuation is greater (0·60 Å as against 0·57 Å). Their lowest frequency of $3{\cdot}1\ cm^{-1}$ is lower than ours, but there is striking agreement in the frequency distributions below $60\ cm^{-1}$: they have 126 such frequencies and we have 123. (For Gō *et al.* (1983), the corresponding number is 91.) The lowest frequency mode of Brooks & Karplus is overall β-hairpin bending and the second lowest is twisting of the loops near the 14–38-disulphide bridge. In our study, these motions also occur in the two lowest modes (in the TE130 conformation, they are in the same order, whereas in the TX conformation the order of the modes is inverted).

The close agreement between our study and that of Brooks & Karplus (1983) is striking when one considers that they use over eight times as many variables (1740/208). Because all three studies of BPTI normal modes use different potentials, it is not possible to estimate the effect of the co-ordinate systems used.

(iii) *Limitations of normal-mode dynamics*

The potential energy function used here is simpler than that used for molecular dynamics and energy minimization calculations (Levitt, 1983*a*,*b*). Hydrogen bonds have no explicit directionality and no (ϕ, ψ) potential is included to compensate for the excluded 1:4 interactions. All solvent molecules (including the four internal water molecules usually included in calculations on BPTI *in vacuo*), are excluded and no effective solvent interaction is used. The more realistic potential used before (Levitt, 1983*a*,*b*) could certainly be used with the present method, but this is not expected to have a major effect. One can include the bound water molecules at a cost of an additional six degrees of freedom per water molecule, but diffusing water molecules cannot be treated as they do not move in a harmonic potential well.

The presence of the surrounding water would be expected to have two major effects: (1) reduce the frequencies by reducing the surface energy of the protein. This energy arises from the surface tension of the protein/vacuum interface and is not expected to change frequencies much. (2) Damp the modes changing the nature of the motion from a harmonic vibration to a diffusive random walk. Because the largest r.m.s. fluctuations in the lowest frequency modes are less than 0·3 Å, i.e. less than the diameter of a water molecule (2 to 3 Å), the macroscopic viscosity of water may not be applicable. It is important to remember that viscous damping can only affect time-dependent behaviour: equilibrium properties like r.m.s. fluctuations, temperature factors and correlation coefficients should be unaffected.

(b) *Nature of protein vibrations*

(i) *Importance of lowest frequency modes*

A few of the lowest frequency modes of vibration (between 3 and 8) dominate the overall motion of the chain in that they account for 70% of the average α-carbon r.m.s. fluctuation. These modes also account for the characteristic distribution of this r.m.s. fluctuation along the chain. Each of these modes involves segmental motion in which up to ten residues move as a rigid body. The nature of the motion in each mode is very characteristic and can be described in terms of bending, twisting and arm-swinging motions. The α-helices tend to move approximately as rigid bodies whereas the β-sheets are deformed by bending and twisting. The very lowest frequency very often corresponds to the dominant motion expected for an isotropic elastic solid of the same shape. For example, the elongated BPTI structure bends like a rod; the larger proteins, ribonuclease and lysozyme, twist or bend to open and close the active-site cleft that cuts across the centre of the globular molecule. The fact that so few modes account for the atomic motion of the backbone suggests that these modes may be useful for analysis of molecular dynamics trajectories and refinement of protein temperature factors in terms of a small number of modes; this work is in progress.

(ii) *Atomic mobility*

The r.m.s. fluctuation of each α-carbon atom depends on its position along the polypeptide chain. This dependence correlates well with the pattern of hydrogen bonds that define the regions of α-helix and β-sheet secondary structure in BPTI protein: the atoms of residues in α-helices and β-sheets vibrate least, while those in exposed loops or the chain termini vibrate most.

(c) *Correlation with experiment*

(i) *Frequencies*

The low-frequency torsional vibrations calculated here have frequencies between 2 and 200 cm^{-1}. Raman spectroscopy can detect frequencies down to about 20 cm^{-1}. Results on lysozyme (Genzel *et al.*, 1976; Peticolas, 1979) show two broad peaks at 75 and at 25 cm^{-1}, with the peak at 25 cm^{-1} showing much greater intensity. For chymotrypsin (Brown *et al.*, 1972; Peticolas, 1979), there is also a peak at 30 cm^{-1}, but the peak at 70 cm^{-1} is very weak. More recently, the low-frequency Raman spectra of a wide range of proteins were compared (Painter *et al.*, 1982). The peaks that were observed ranged from 14 to 36 cm^{-1} and were not correlated with protein size. It is of interest that our results show that torsion vibrations in β-sheets occur between 15 and 40 cm^{-1}, whereas those in α-helices occur between 55 and 115 cm^{-1}. It is tempting to suggest that the Raman band at 75 cm^{-1} is due to α-helix while that at 25 cm^{-1} is due to β-sheet. The greater proportion of β-sheet in chymotrypsin (7% α, 55% β) compared to lysozyme (46% α, 19% β) (Levitt & Greer, 1977) would explain the relative intensities of the 25 cm^{-1} and 75 cm^{-1} bands in the Raman spectra of these two proteins. Interestingly, the neutron scattering peaks of α-helical polyalanine at 80 cm^{-1} and β-sheet polyalanine at 50 cm^{-1} (Drexel & Peticolas, 1975) partly support this suggestion.

Another method sensitive to frequencies below 100 cm^{-1} is neutron time-of-flight spectroscopy. For lysozyme, recent results show peaks at 75 and 25 cm^{-1}, with the peak at 75 cm^{-1} being of greater intensity (Bartunik *et al.*, 1982). Unfortunately, neither of these two methods is able to provide information about the frequency range below 10 cm^{-1} that seems most important for low-frequency collective vibrations; we hope that the present results will encourage such experimental work.

(ii) *Temperature factors*

The agreement between the atomic temperature factors calculated here and the new refined values obtained experimentally is good (correlation coefficient of about 0·6). This correlation is striking when one considers that the calculated values

ignore unit-cell packing interactions and that the experimental values depend on both static disorder and atomic motion. A full comparison with experiment, now in progress, will include the anisotropic motions of all atoms and allow for static crystal disorder and overall rigid-body motion in the unit cell. It is of interest to note that the calculated temperature factors are not sensitive to the exact atomic co-ordinates; the values for TX and TE130 are almost identical although these conformations differ by a 1·8 Å r.m.s. deviation of the α-carbon atoms. This suggests that the present method could be used to calculate temperature factors from the initial unrefined protein co-ordinates and that these factors could be of use in the subsequent refinement of the co-ordinates against the X-ray data.

5. Conclusions

This paper has introduced, described in detail and verified a new method for calculating the normal modes of vibration of biological macromolecules. Application of the method to four proteins provides a concise, complete description of protein dynamics. The chain motion is shown to be collective and dominated by segmental motion in the low frequency modes. Local torsion angle motion occurs over a wide range of frequencies although there are characteristic frequencies for residues in α-helix or β-sheet secondary structure. The correlation matrices for all pairs of α-carbons and single-bond torsion angles show that while pairs of α-carbon atoms closer together than 20 Å show significant correlation, only pairs of torsion angles in those residues that are adjacent in sequence or connected by hydrogen or disulphide bonds show significant correlation.

Although the results given here agree with experiments on both the frequency of motion and the extent of this motion, it is necessary to consider the functional relevance of the normal modes of vibration. An important aspect of the way protein molecules function as enzymes involves changes of conformation and transmission of these changes across the structure. Conformational rearrangements require transition from the energy minimum of the first state over an energy barrier to the energy minimum of the second state. This process cannot be reproduced by a normal-mode calculation that can only calculate the properties of each state individually. Normal modes are, however, useful for the following reasons.

(1) The free energies of all states can be calculated (see Table 1) and used to estimate rates of processes.

(2) The initial conformational change will probably be a superposition of a few of the lowest frequency modes, as this allows the largest motions for the smallest increase in energy.

(3) The long range of the α-carbon correlations suggests a way to transmit conformational changes; binding of a substrate could affect a low-frequency collective vibration that would affect the motion of atoms some distance away.

One of us (C.S.) is grateful to the Minerva and EMBO fellowship programs and to the Deutsche Forschungsgemeinschaft project 'Protein Structure Theory' for support. We are grateful to the Weizmann Institute for providing ample computing and graphics facilities and to the European Molecular Biology Laboratory for use of colour vector graphics equipment.

References

Bartunik, H. D., Jolles, P., Berthou, J. & Dianoux, A. (1982). *Biopolymers,* **21,** 43–50.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1965). *Nature (London),* **206**, 757–761.

Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). *Proc. Roy. Soc. ser. B,* **167**, 365–377.

Brooks, B. & Karplus, M. (1983). *Proc. Nat. Acad. Sci. U.S.A.* **80**, 6571–6575.

Brown, K. G., Erfurth, S. C., Small, E. W. & Peticolas, W. L. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 1467–1469.

Chothia, C. & Lesk, A. M. (1984). *J. Mol. Biol.* **174**, 175–191.

Deisenhofer, J. & Steigemann, W. (1975). *Acta Crystallogr. sect. B,* **31**, 238–250.

Drexel, W. & Peticolas, W. (1975). *Biopolymers,* **14**, 715–721.

Fletcher, R. (1972). U.K. Atomic Energy Authority Harwell, Report AERE R7125.

Genzel, L., Keilmann, F., Martin, T. P., Winterling, G., Yacoby, Y., Fröhlich, H. & Makinen, M. W. (1976). *Biopolymers,* **15**, 219–225.

Gō, M. & Gō, N. (1976). *Biopolymers,* **15**, 1119–1127.

Gō, N., Noguti, T. & Nishikawa (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 3696–3700.

Goldstein, H. (1950). *Classical Mechanics,* Addison-Wesley, Reading, Massachusetts.

Hagler, A. T., Stern, P. S., Sharon, R., Becker, J. M. & Naider, F. (1979). *J. Amer. Chem. Soc.* **101**, 6842–6852.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London),* **290**, 107–113.

Huber, R. & Bennett, W. S. (1983). *Biopolymers,* **22**, 261–279.

Huber, R., Kukla, D., Rühlmann, A. & Steigemann, W. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 141–150.

Itoh, K. & Shimanouchi, T. (1970). *Biopolymers,* **9**, 383–399.

Kabsch, W. (1976). *Acta Crystallogr. sect. A,* **32**, 922–923.

Kabsch, W. & Sander, C. (1983). *Biopolymers,* **22**, 2577–2637.

Karplus, M. & McCammon, J. A. (1981). *CRC Crit. Rev. Biochem.* **9**, 293–349.

Katz, H., Walter, R. & Somorjai, R. L. (1979). *Computers & Chemistry,* **3**, 25–32.

Kubo, R. (1967). *Statistical Mechanics,* p. 178, North-Holland, Amsterdam.

Levitt, M. (1972). Ph.D. thesis, University of Cambridge.

Levitt, M. (1983*a*). *J. Mol. Biol.* **168**, 595–620.

Levitt, M. (1983*b*). *J. Mol. Biol.* **168**, 621–657.

Levitt, M. (1983*c*). *J. Mol. Biol.* **170**, 723–764.

Levitt, M. & Greer, J. (1977). *J. Mol. Biol.* **114**, 181–239.

Levitt, M., Sander, C. & Stern, P. S. (1983). *Int. J. Quant. Chem.: Quant. Biol. Symp.* **10**, 181–199.
Levy, R. & Karplus, M. (1979). *Biopolymers,* **18**, 2465–2495.
Lifson, S. & Stern, P. S. (1982). *J. Chem. Phys.* **77**, 4542–4550.
Lifson, S. & Warshel, A. (1968). *J. Chem. Phys.* **49**, 5116–5129.
McLachlan, A. D. (1979). *J. Mol. Biol.* **128**, 49–79.
Noguti, T. & Gō, N. (1982). *Nature (London),* **296**, 776–778.
Noguti, T. & Gō, N. (1983*a*). *J. Phys. Soc. Japan,* **52**, 3283–3288.
Noguti, T. & Gō, N. (1983*b*). *J. Phys. Soc. Japan,* **52**, 3685–3690.
Painter, P. C., Mosher, L. E. & Rhoads, C. (1982). *Biopolymers,* **21**, 1469–1472.
Peticolas, W. (1979). In *Methods in Enzymology* (Hirs, C. H. W. & Timasheff, S. N., eds), **61**, 425–458.
Remington, S. J., Wiegand, G. & Huber, R. (1982). *J. Mol. Biol.* **158**, 111–152.
Richards, F. H. & Wyckoff, H. W. (1971). In *The Enzymes* (Boyer, P. D., ed.), 3rd edit., vol. 4, pp. 647–806, Academic Press, New York.
Sander, C. & Stern, P. S. (1979). In *Rapport d'Activité Scientifique du CECAM,* Workshop du CECAM: Analyse des Structures des Proteines, pp. 121–127, Université Paris-Sud, Orsay.
Shimanouchi, T. (1970). In *Physical Chemistry* (Eyring, H., Henderson, D. & Jost, W., eds), vol. 4, chap. 6, Academic Press, New York.
Sternberg, M. J. E., Grace, D. E. P. & Phillips, D. C. (1979). *J. Mol. Biol.* **130**, 231–253.
Tasumi, M., Takeuchi, H., Ataka, S., Dwivedi, A. M. & Krimm, S. (1982). *Biopolymers,* **21**, 711–714.
Twardowski, J. (1978). *Biopolymers,* **17**, 181–190.
van Gunsteren, W. F., Berendsen, H. J. C., Hermans, J., Hol, W. G. J. & Postma, J. P. M. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 4315–4319.
Walter, J. & Huber, R. (1983). *J. Mol. Biol.* **167**, 911–917.
Wetlaufer, D. B. (1973). *Proc. Nat. Acad. Sci., U.S.A.* **70**, 697–701.
Wilkinson, J. H. & Reinsch, C. (1971). *Handbook for Automatic Computation,* vol. 2. *Linear Algebra,* Springer-Verlag, Berlin.
Wilson, E. B., Decius, J. C. & Cross, P. C. (1955). *Molecular Vibrations,* McGraw-Hill, New York.
Wlodawer, A. & Sjölin, L. (1983). *Biochemistry,* **22**, 2720–2728.

Edited by J. C. Kendrew