ANTHONY ARTALE
MED LIB NATHAN CUMMINGS CTR (S-46)
MEMORIAL SLOAN KETTERING CANCER CTR
1275 YORK AVENUE
NEW YORK, NY 10021
UNITED STATES



**Title:** **INTERNATIONAL JOURNAL OF QUANTUM CHEMISTRY. PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM ON QUANTUM BIOLOGY AND QUANTUM PHARMACOLOGY.**

DB Ref. No.: IRN10196006
ISSN: ISSN03608832
Vol./Issue: 10
Date: 1983
Pages: 181-199
Article Title: THE NORMAL MODES OF A PROTEIN: NATIVE BOVINE PANCREATIC TRYPSIN
Article Author: LEVITT M, STERN P
Report Number: IRN10196006
Publisher: WILEY,

**INSTRUCTIONS: PATRON REQUESTS COLOUR IF AVAILABLE. THANK YOU.**

**CISTI will be closed from December 23, 2005 at noon (ET) to January 3, 2006 at 8:30 am (ET).**

**Estimated cost for this 19 page document: $10.2 document supply fee + $30 copyright = $40.2**

# The Normal Modes of a Protein: Native Bovine Pancreatic Trypsin Inhibitor

MICHAEL LEVITT, CHRISTIAN SANDER,* AND PETER S. STERN
*Chemical Physics Department, Weizmann Institute of Science, 76100 Rehovot, Israel*

## Abstract

A completely general treatment of normal mode analysis is developed that can be used with any potential energy function and any set of generalized coordinates. The method is applied to the calculation of the normal modes of the small protein bovine pancreatic trypsin inhibitor that has been the subject of many previous theoretical studies. The potential energy function used comprises a torsion angle potential, a van der Waals potential between nonbonded pairs of atoms, and a hydrogen bond potential. Therefore, the generalized coordinates used are the 208 $\phi$, $\psi$, and $\chi$ torsion angles about single bonds. This eliminates the difficulties inherent in using internal or Cartesian coordinates for a large molecule. Many dynamic properties of the protein may now be calculated in the normal mode description. In particular, the rms magnitudes and pair correlations of the fluctuations in positions and velocities of the $\alpha$-carbon atoms and various classes of torsion angles, such as backbone, side chain, $\beta$-sheet, and $\alpha$-helix, are calculated and analyzed to identify the most correlated modes. In addition, the ir intensities are calculated.

## Introduction

Protein flexibility has been the subject of many experimental and theoretical investigations in recent years as is well documented in several reviews [1–7]. This interest stems, in part, from the realization that dynamics, as well as structure, is an essential aspect of protein activity. Protein flexibility plays an important role in a number of biological functions such as enzyme catalysis [8], hemoglobin cooperativity [9], immunoglobin action [10], and the allosteric effect [11], to name but a few.

On the experimental side, protein dynamics has been investigated by several techniques. These include ir [12] and Raman [13–15] spectroscopy, analysis of X-ray temperature factors [16–18], fluorescence depolarization [19] and quenching [20], and NMR [6,21–25]. Theoretical calculations on protein dynamics include determination and analysis of an enzymatic reaction path using quantum mechanical and empirical potential functions [26,27], vibrational analysis of proteins [28–32] and their secondary structure elements [33–35], and molecular dynamics calculations [36–38]. For further details, the reader is referred to the excellent and extensive review by Karplus and McCammon [3].

*Present address: Department of Biophysics, Max Planck Institute of Medical Research, D-6900 Heidelberg, Germany.

International Journal of Quantum Chemistry: Quantum Biology Symposium 10, 181–199 (1983)

CCC 0360-8832/83/010181-19$04.00

In this report, the problem of protein dynamics is addressed from the point of view of vibrational flexibility. Thus, the normal modes of vibration of the globular protein bovine pancreatic trypsin inhibitor (BPTI), in its native state, are calculated and analyzed. Solving for the normal modes in Cartesian [39] or internal coordinates [40] requires diagonalization of a $3N \times 3N$ matrix, where $N$ is the number of atoms in the protein. This is clearly not feasible even for as small a protein as BPTI, with $N = 515$ (neglecting all but the amide hydrogen atoms). Therefore, the problem is treated as a function of only the $\phi$, $\psi$, and $\chi$ torsion angles about single bonds [41], of which there are but 208. As a result, only 208 eigenvalues of the protein are obtained in this approximation, but the problem becomes tractable and applicable to bigger proteins.

This method has certain theoretical limitations. Firstly, it approximates the vibrational displacements of the atoms from their equilibrium positions by a sum of quadratic terms. This, of course, neglects the possible anharmonicity of the actual displacements. Secondly, it imposes an artificial rigidity upon the protein molecule, since it does not permit relaxation of bonds and, more important, bond angles. Inclusion of these degrees of freedom would reduce the steepness of potential barriers [32,42,43] and, in general, reduce frequencies by softening the protein normal modes.

On the other hand, it is a valid approximation which permits the application of the normal mode approach to systems heretofore considered intractable. Considering the fact that this approach requires the additional approximation that the molecule be treated *in vacuo,* thereby neglecting any damping effects of, or energetic interactions with, the solvent, the neglect of anharmonicity and of flexible bonds and bond angles may not be unreasonable. Compared to molecular dynamics calculations which require on the order of 100 times more computer time, the method developed here for the study of protein dynamics yields all the basic time scales (from $\sim 1.5 \times 10^{-13}$ to $10^{-11}$) at once and provides a description of protein dynamics in terms of the much more comprehensible normal mode description.

## Methods

A completely general formalism is presented for calculating the normal modes of a system using any potential energy function of position, in any set of generalized coordinates, even in the case where these are fewer than the $3N$ degrees of freedom of the $N$-particle system [44]. This method is applied to the small protein BPTI that has been the subject of numerous theoretical [31,36–38,41,45] and experimental [22,23,25,46,47] studies. Here, the generalized coordinates are chosen to be the 208 $\phi$, $\psi$, and $\chi$ torsion angles about single bonds plus the six coordinates defining the center of mass and the angles of rotation of the undistorted molecule. These latter six coordinates must be fixed such that varying a set of torsion angles does not result in an arbitrary rotation of the molecule or translation of its center of mass. The steps in the calculation are as follows:

n dynamics is addressed from the point
s, the normal modes of vibration of the
ɔsin inhibitor (BPTI), in its native state,
ɔr the normal modes in Cartesian [39] or
ɔnalization of a $3N \times 3N$ matrix, where
ı. This is clearly not feasible even for as
5 (neglecting all but the amide hydrogen
ed as a function of only the $\phi$, $\psi$, and $\chi$
of which there are but 208. As a result,
e obtained in this approximation, but the
ble to bigger proteins.
limitations. Firstly, it approximates the
from their equilibrium positions by a sum
eglects the possible anharmonicity of the
oses an artificial rigidity upon the protein
laxation of bonds and, more important,
s of freedom would reduce the steepness
general, reduce frequencies by softening

roximation which permits the application
ns heretofore considered intractable. Con-
equires the additional approximation that
eby neglecting any damping effects of, or
nt, the neglect of anharmonicity and of
t be unreasonable. Compared to molecular
ɔn the order of 100 times more computer
ıe study of protein dynamics yields all the
to $10^{-11}$) at once and provides a description
nuch more comprehensible normal mode

## ethods

ɔresented for calculating the normal modes
y function of position, in any set of gen-
where these are fewer than the $3N$ degrees
[44]. This method is applied to the small
of numerous theoretical [31,36–38,41,45]
tudies. Here, the generalized coordinates
torsion angles about single bonds plus the
f mass and the angles of rotation of the
coordinates must be fixed such that varying
in an arbitrary rotation of the molecule or
steps in the calculation are as follows:



1. Definition of the Cartesian coordinates $\mathbf{r}_k$, in terms of the generalized coordinates $q_i$, and minimization of the potential energy with respect to the coordinates $q_i$. The potential energy function used comprises torsional, van der Waals, and hydrogen bond terms of the form [1,45]

$$V = \sum_{\substack{\text{torsion}\\ \text{angles}}} K_\phi [1 - \cos(n\phi + \delta)] + \sum_{\substack{\text{nonbonded}\\ \text{pairs}}} \varepsilon\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right] + \sum_{\substack{\text{hydrogen}\\ \text{bond pairs}}} f_{HB}(r,\theta).$$

This function, the scheme for calculating first derivatives, and the minimization algorithm are those used in a previous study of protein folding [45] and are described there in detail. It should be emphasized that for the calculation of the normal modes to be correct, the conformation used must be at a precise minimum [44]. This is achieved by using double-precision arithmetic (i.e., 16 significant figures) and the convergent VA09D minimization algorithm from the Harwell subroutine library [48]. The largest energy derivatives after 475 iteration are $4 \times 10^{-4}$ kcal/mol rad.

2. Calculation of **V** and **T.** The matrix **V,** which contains the second derivatives of the potential energy with respect to the generalized coordinates $q_i$, is given by

$$V_{ij} = \left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right)_{q = q^0}.$$

$V$ can be calculated analytically [49–51], but it was easier to use numerical differentation of $\partial V/\partial q_i$ to give

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

The matrix **T,** which contains the second derivatives of the kinetic energy $T$, with respect to the generalized velocities $q_i$ is given by

$$T_{ij} = \left(\frac{\partial^2 T}{\partial \dot{q}_i \partial \dot{q}_j}\right)_{q = q^0}.$$

For a system of $N$ atoms with masses $m_k$ and Cartesian coordinates $\mathbf{r}_k$, the kinetic energy is given by

$$T = \frac{1}{2}\sum_k^N m_k \dot{\mathbf{r}}_k^2. \tag{1}$$

Small changes in the generalized coordination $\delta q_i$ cause small changes in the Cartesian coordinates that are given by

$$\delta \mathbf{r}_k = \sum_i^N \left(\frac{\partial \mathbf{r}_k}{\partial q_i}\right)\delta q_i.$$

If these changes occur in time $\delta t$, then

$$\frac{\delta \mathbf{r}_k}{\delta t} = \sum_i^N \left(\frac{\partial \mathbf{r}_k}{\partial q_i}\right) \frac{\delta q_i}{\delta t}$$

or

$$\dot{\mathbf{r}}_k = \sum_i^N \left(\frac{\partial \mathbf{r}_k}{\partial q_i}\right) \dot{q}_i.$$

Substituting for $\dot{\mathbf{r}}_k$ in eq. (1) gives

$$\begin{aligned} T &= \frac{1}{2} \sum_k^N m_k \sum_i^n \frac{\partial \mathbf{r}_k}{\partial q_i} \dot{q}_i \sum_j^n \frac{\partial \mathbf{r}_k}{\partial q_j} \dot{q}_j \\ &= \frac{1}{2} \sum_i^n \sum_j^n \sum_k^N m_k \left[ \frac{\partial \mathbf{r}_k}{\partial q_i} \cdot \frac{\partial \mathbf{r}_k}{\partial q_j} \right] \dot{q}_i \, \dot{q}_j. \end{aligned} \tag{2}$$

In the generalized coordinates $q_i$, the kinetic energy can be written as a quadratic function of the velocities $\dot{q}_i$,

$$T = \frac{1}{2} \sum_{ij}^n T_{ij} \dot{q}_i \dot{q}_j.$$

Comparison with eq. (2) gives

$$T_{ij} = \sum_k^N m_k \left( \frac{\partial \mathbf{r}_k}{\partial q_i} \cdot \frac{\partial \mathbf{r}_k}{\partial q_j} \right).$$

The calculation of $(\partial V/\partial q_i)_{q = q^0}$ presents no problems since the potential energy is not a function of translation or rotation of the molecule. However, the kinetic energy must not include any translation or rotation of the system, so that special care must be taken in calculating $\partial \mathbf{r}_k/\partial q_i$. In the equilibrium conformation, one of the torsion angles $q_i$ is changed to $q_i = q_i^0 + \delta q_i$. This results in a change in the Cartesian coordinates of many atoms from $\mathbf{r}_k^0$ to $\mathbf{r}_k$ possibly resulting in a change in the center of mass. Thus, the new coordinates $\mathbf{r}_k$ must be superimposed as a rigid body onto the initial coordinates $\mathbf{r}_k^0$. Use of transformational matrices [52,53] calculated with Cartesian coordinates multiplied by $\mathbf{m}_k^{1/2}$ eliminates any rigid body rotation or translation of the molecule and gives a new set of Cartesian coordinates $\mathbf{r}_k'$. The matrix $\partial \mathbf{r}_k/\partial q_i$ is now calculated numerically as $(\mathbf{r}_k' - \mathbf{r}_k^0)/\delta q_i$ for all $q_i$. The transformation matrix $\partial \mathbf{r}_k/\partial q_i$ is used below to calculate molecular properties from the normal coordinates.

(3) Solution of the equations of motion to get the eigenvalues $\mathbf{\Lambda}$ and the eigenvectors $\mathbf{A}_k$, and expression of the generalized coordinates $q_i$ in terms of the normal coordinates $Q_k$. The equations of motion for the system in terms of the generalized coordinates $q_i$ are [44]

$$\left(\frac{\partial \mathbf{r}_k}{\partial q_i}\right) \frac{\delta q_i}{\delta t}$$

$$\left(\frac{\partial \mathbf{r}_k}{\partial q_i}\right) \dot{q}_i.$$

$$\frac{\mathbf{r}_k}{q_i} \dot{q}_i \sum_j^n \frac{\partial \mathbf{r}_k}{\partial q_j} \dot{q}_j \qquad (2)$$

$$n_k \left[\frac{\partial \mathbf{r}_k}{\partial q_i} \cdot \frac{\partial \mathbf{r}_k}{\partial q_j}\right] \dot{q}_i \dot{q}_j.$$

netic energy can be written as a quadratic

$$\sum_{ij}^n T_{ij} \dot{q}_i \dot{q}_j.$$

$$\left(\frac{\partial \mathbf{r}_k}{\partial q_i} \cdot \frac{\partial \mathbf{r}_k}{\partial q_j}\right).$$

its no problems since the potential energy on of the molecule. However, the kinetic n or rotation of the system, so that special $q_i$. In the equilibrium conformation, one $q_i = q_i^0 + \delta q_i$. This results in a change atoms from $\mathbf{r}_k^0$ to $\mathbf{r}_k$ possibly resulting in , the new coordinates $\mathbf{r}_k$ must be super- l coordinates $\mathbf{r}_k^0$. Use of transformational sian coordinates multiplied by $\mathbf{m}_k^{1/2}$ elim- ation of the molecule and gives a new set x $\partial \mathbf{r}_k/\partial q_i$ is now calculated numerically as rmation matrix $\partial \mathbf{r}_k/\partial q_i$ is used below to e normal coordinates.

notion to get the eigenvalues $\mathbf{\Lambda}$ and the generalized coordinates $q_i$ in terms of the of motion for the system in terms of the



$$V = \frac{1}{2} \sum_{i,j}^n V_{ij}(q_i - q_i^0)(q_j - q_j^0) = \frac{1}{2}(\mathbf{q} - \mathbf{q}^0)^T \mathbf{V}(\mathbf{q} - \mathbf{q}^0),$$

$$T = \frac{1}{2} \sum_{ij}^n T_{ij} \dot{q}_i \dot{q}_j = \frac{1}{2} \dot{\mathbf{q}}^T \mathbf{T} \dot{\mathbf{q}},$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \left(\frac{\partial L}{\partial q_i}\right).$$

The Lagrangian $L = T - V$; and from the first two equations, we have $\partial L/\partial \dot{q}_i = \sum_j^n T_{ij} \dot{q}_j$ and $\partial L/\partial q_i = -\sum_j^n V_{ij}(q_j - q_j^0)$, and Lagrange's equation becomes

$$\sum_j^n T_{ij} \ddot{q}_j - \sum_j^n V_{ij}(q_j - q_j^0) = 0, \qquad (3)$$

which has solutions of the form

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

Substitution of $q_j$ and $\ddot{q}_j$ into eq. (3) ultimately gives the equations of motion written in matrix notation as

$$\mathbf{VA} = \mathbf{TA\Lambda}, \qquad (5)$$

where $\mathbf{\Lambda}$ is diagonal and $\Lambda_{ii} = \omega_i^2$. Defining $Q_k = \alpha_k \cos(\omega_k + \delta_k)$ gives, from eq. (4),

$$q_j - q_j^0 = \sum_k^n A_{jk} Q_k \qquad \text{or} \qquad \mathbf{q} - \mathbf{q}^0 = \mathbf{AQ}$$

as the relationship between the coordinates $\mathbf{q}$ and $\mathbf{Q}$. If $\mathbf{A}$ is chosen so that $\mathbf{A}^T\mathbf{TA} = \mathbf{1}$, then

$$T = \frac{1}{2} \dot{\mathbf{q}} \mathbf{T} \dot{\mathbf{q}} = \frac{1}{2} \dot{\mathbf{Q}}^T \mathbf{A}^T \mathbf{T} \mathbf{A} \mathbf{Q} = \frac{1}{2} \dot{\mathbf{Q}}^T \dot{\mathbf{Q}} = \frac{1}{2} \sum_i^n \dot{Q}^2.$$

Multiplying the equations of motion (5) on the left by $\mathbf{A}^T$ gives $\mathbf{A}^T\mathbf{VA} = \mathbf{A}^T\mathbf{TA\Lambda} = \mathbf{\Lambda}$, so that

$$V = \frac{1}{2}(\mathbf{q} - \mathbf{q}^0)^T \mathbf{V}(\mathbf{q} - \mathbf{q}^0) = \frac{1}{2} \mathbf{Q}^T \mathbf{A}^T \mathbf{V} \mathbf{A} \mathbf{Q}$$

$$= \frac{1}{2} \mathbf{Q}^T \mathbf{\Lambda} \mathbf{Q} = \frac{1}{2} \sum_i^n \Lambda_{ii} Q_i^2 = \frac{1}{2} \sum_i^n \omega_i^2 Q_i^2.$$

Because the potential and kinetic energies are simple sums of squares of $Q_i$ and $\dot{Q}_i$, the coordinates $Q_i$ are normal coordinates. From the expression for $T$, it is seen that they are also mass scaled. Solution of eq. (5) for $\mathbf{A}$ and $\mathbf{\Lambda}$ with the

normalization condition $\mathbf{A}^T\mathbf{T}\mathbf{A} = \mathbf{1}$ is accomplished by standard methods (for example, subroutine F02AEF from the Numerical Algorithms Group library [54]). Equation (4) gives the complete dynamic behavior of the system where, for the *kth* normal mode, the angular frequency $\omega_k = \Lambda_{kk}^{1/2}$, and the amplitude $\alpha_k$ and the phase $\delta_k$ depend upon the positions and velocities at time $t = 0$.

*Calculation of Time-Averaged Properties*

Time-averaged properties of the motion depend only upon the amplitude $\alpha_k$. For example, the correlation coefficient of a pair of coordinates is given by

$$\begin{aligned}\langle \Delta q_i(\tau + t)\Delta q_j(\tau)\rangle &= \left\langle \sum_k^n A_{ik}Q_k(\tau + t) \sum_l^n A_{jl}Q_l(\tau)\right\rangle \\ &= \sum_k^n \sum_l^n A_{ik}A_{jl}\langle Q_k(\tau + t)Q_l(\tau)\rangle \\ &= \frac{1}{2}\sum_k^n \sum_l^n A_{ik}A_{jl}\delta_{kl}\alpha_k^2 \cos \omega_k t \\ &= \frac{1}{2}\sum_k^n \sum_l^n A_{ik}A_{jk}\alpha_k^2 \cos \omega_k t.\end{aligned}$$

At time $t = 0$, of course, the $\cos \omega_k t$ term is equal to 1. The expression for the velocity correlations is obtained similarly by taking the time derivatives of eq. (4),

$$\langle \Delta\dot{q}_i(\tau + t)\Delta\dot{q}_j(\tau)\rangle = \frac{1}{2}\sum_k^n A_{ik}A_{jk}\alpha_k^2\omega_k^2 \cos \omega_k t.$$

Consider some parameter $\Delta p_i$ of this system that is a linear function of the change in generalized coordinates $\Delta q_k$, i.e., $\Delta p_i = \sum_k^n P_{ik}\,\Delta q_k$. The correlation coefficients of the $\Delta p_i$ are

$$\begin{aligned}\langle \Delta p_i(\tau)\Delta p_j(\tau)\rangle &= \sum_k^n \sum_k^n P_{ik}P_{jl}\langle \Delta q_k(\tau)\Delta q_l(\tau)\rangle \\ &= \frac{1}{2}\sum_k^n \sum_l^n \sum_m^n P_{ik}P_{jl}A_{km}A_{lm}\alpha_m^2.\end{aligned}$$

For example, the change in Cartesian coordinates is given by $\mathbf{\Delta r} = \mathbf{P\Delta q}$, where $P_{ik} = \partial \mathbf{r}_i/\partial q_k$. The change in the dipole moment $\boldsymbol{\mu}$, is used to calculate the ir intensities:

$$\begin{aligned}\boldsymbol{\mu} &= \boldsymbol{\mu}^0 + \sum_k^n \frac{\partial \boldsymbol{\mu}}{\partial Q_k} + \text{higher terms} \\ &= \boldsymbol{\mu}^0 + \sum_j^n \sum_k^n \frac{\partial \boldsymbol{\mu}}{\partial q_j} A_{jk},\end{aligned}$$

where $\partial\boldsymbol{\mu}/\partial q_j = \sum_i^N e_i \partial \mathbf{r}_i/\partial q_j$, and either empirical force field or quantum mechanical values can be used for the partial atomic charges $e_i$.

*Thermal Amplitudes and Quantum Corrections*

Quantum mechanically, each mode will behave like an harmonic oscillator with energy levels $\hbar\omega$ apart. The mean square fluctuation of the normal mode variable $Q_i$ is related to the temperature [55] by

$$\langle Q_i^2(\tau)\rangle = \frac{\hbar}{2\omega_i}\coth\left(\frac{\hbar\omega_i}{2k_BT}\right) = \frac{1}{2}\alpha_i^2$$

since for $y$,

$$\lim_{y\to 0}\coth(y) \equiv \lim_{y\to 0}\frac{(e^y + e^{-y})}{(e^y - e^{-y})} = \frac{1}{y}$$

and

$$\langle Q_i^2(\tau)\rangle = \frac{\hbar}{2\omega_i}\frac{2k_BT}{\hbar\omega_i} = \frac{k_BT}{\omega_i^2} = \frac{\alpha_i^2}{2}.$$

Thus, $\alpha_i^2 = 2k_BT/\omega_i^2$ and $\langle V\rangle = 1/2\omega_i^2\,\langle Q_i^2(\tau)\rangle = 1/2k_BT$, which is the classical value. At 310 K, $y = \nu_i/430$ for the frequency $\nu_i$ expressed in $cm^{-1}$. Because $\nu_i$ is much less than 200 $cm^{-1}$ for most of the protein normal modes calculated here, the quantum mechanical result is almost identical to the classical result.

*Computing Considerations*

One of the major advantages of the normal mode analysis of protein dynamics is that requirements of computer resources are much smaller than with dynamic simulations. For BPTI protein, which has 515 atoms and 208 single-bond torsion angles, one evaluation of the potential energy and its analytical first derivatives takes 2.2 s (all times are cpu time for the IBM 370/165 computer that is rated at about 2.5 Mips, or million instructions per second). Reaching the potential energy minimum takes 14 min, calculating the **V** matrix numerically takes 7 min, calculating the **T** matrix and solving the equations of motion takes 4 min, and analysis as presented here takes 2 min. Thus, the complete normal mode treatment takes 27 min. For a protein twice as large as BPTI (for example, lysozyme), the time requirement is expected to be four times more (about 2 h).

Memory requirements are perhaps more severe. Minimizing the energy and calculating the **V** matrix requires 500 kbytes of memory on the IBM 370/165 with an extensively overlayed program. Calculating the **T** matrix and the analysis requires 1500 kbytes and is done on an IBM 4341-2 computer that uses virtual memory. For lysozyme, 6000 kbytes of virtual memory would be required.

The major difficulty of using the present methods on a large protein are concerned with comprehensible presentation of the results. Even for BPTI, there

are 208 normal modes, and each mode involves changes in all 208 single-bond torsion angles.

## Results and Discussion

Solution of the equations of motion, eq. (5), gives the fundamental frequencies of vibration of the protein molecule in torsional space. Thus, we present in Figure 1(A) a histogram of the number of modes for the 208 low-frequency torsional modes of BPTI. Although calculation of the allowed Raman frequencies has not been carried out, the overall shape of the histogram is similar to those of the Raman spectra of protein crystals [13–15]. The peak centers at about 30–40 $cm^{-1}$, which is about 10 $cm^{-1}$ higher than the Raman result for protein crystals other than BPTI. Anharmonicity and intermolecular crystal forces might well account for the difference, but so might the lower mass of BPTI. In Figure 1(B), the relative ir intensities are shown. They were calculated as described above, using consistent force field partial charges [56]. Use of *ab initio* partial charges derived using a 4-31*G* basis set [57] gives virtually identical results. While the Raman spectra are expected to give broad bands, one expects to find few active ir peaks below 50 $cm^{-1}$.

Figure 2 compares the rms α-carbon fluctuations with the rms torsion angle fluctuations as a function of frequency. Fluctuations in the α-carbon backbone fall off rapidly with frequency. This indicates that the lower modes are likely to be the most important for describing collective protein motion. On the other hand, there are significant torsional displacements throughout the frequency range. This is partially a result of the fact that the backbone torsional angles $\phi$ and $\psi$ are correlated in such a way as to reduce displacements of the backbone

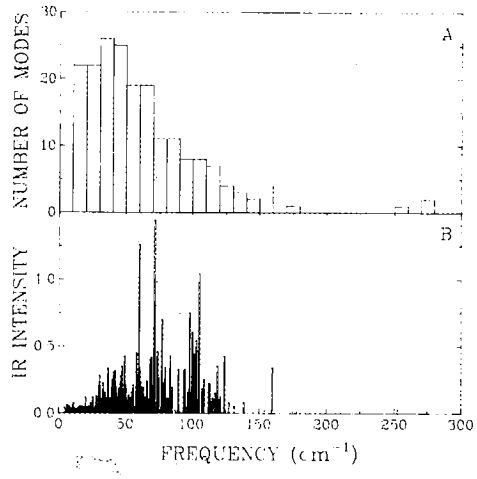


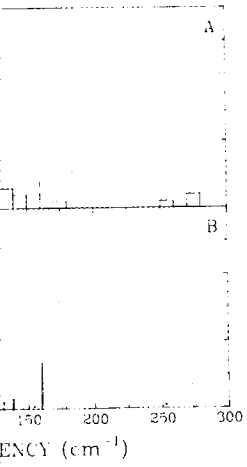

Figure 1. (A) Histogram of the number of modes as a function of frequency. The shape is similar to those of the Raman spectra of protein crystals. (B) Relative ir intensities of the low-frequency modes.

involves changes in all 208 single-bond

## Discussion

q. (5), gives the fundamental frequencies
n torsional space. Thus, we present in
r of modes for the 208 low-frequency
ılation of the allowed Raman frequencies
ape of the histogram is similar to those
als [13–15]. The peak centers at about
igher than the Raman result for protein
y and intermolecular crystal forces might
night the lower mass of BPTI. In Figure
wn. They were calculated as described
ial charges [56]. Use of *ab initio* partial
et [57] gives virtually identical results.
to give broad bands, one expects to find

fluctuations with the rms torsion angle
. Fluctuations in the α-carbon backbone
ndicates that the lower modes are likely
collective protein motion. On the other
displacements throughout the frequency
fact that the backbone torsional angles ϕ
to reduce displacements of the backbone




modes as a function of frequency. The shape is
otein crystals. (B) Relative ir intensities of the



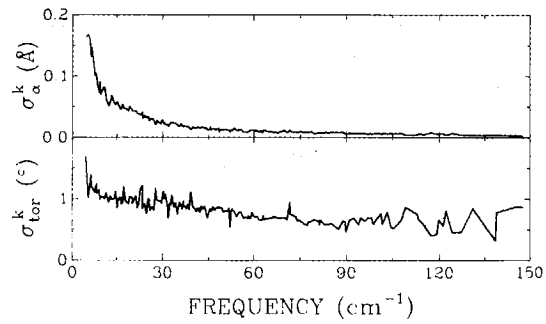



Figure 2. Comparison of the rms $C^\alpha$ fluctuations with the rms torsion angle fluctuations as function of frequency. Fluctuations in the protein backbone fall off rapidly with frequency, while there are significant torsional displacements throughout the frequency range.

atoms, as well as of the fact that side-chain χ angles contribute more to the higher frequency modes.

The overall motion of the backbone, as represented by the α-carbons, is shown in Figure 3. The α-helices and β-sheets are underlined in the figure and clearly correspond to the minima, while the maxima occur at the ends and in loops. These are also the areas most accessible to solvent [38]. Direct comparison of these results with experiment is difficult. NMR experiments on BPTI do find motion of the α-carbons on the time scale seen here [25] but do not identify specific residues or absolute amplitudes. X-ray temperature factors give 0.68 Å for the rms α-carbon fluctuations [3], but these include other contributions to the thermal parameters. Molecular dynamics simulations give 0.60 Å [37,38], while this approach gives 0.56 Å, in good agreement with all the above results. The general features of Figure 3 compare quite well with a similar picture derived from the molecular dynamics work [36,38]. Analysis of the X-ray temperature

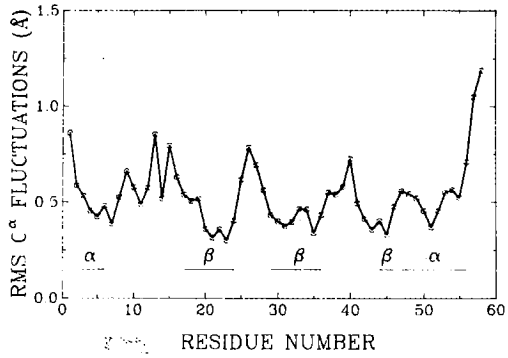



Figure 3. Rms $C^\alpha$ fluctuations (summed over all frequencies) for each residue. The minima are well correlated with secondary structure elements, while the maxima are at the ends and in the connecting loops.

factors of myoglobin [16], lysozyme [17], and ferrocytochrome *c* [18] all yield similar results.

Figure 4 shows the pair correlations of α-carbon atoms plotted against the corresponding α-carbon distances. The superimposed "eyeball" best-fit curve can be interpreted as follows: Residues close to one another (<10 Å apart), such as those that make up a unit of secondary structure (e.g., α-helix or β-sheet), are strongly correlated. At the intermediate distances (~10–25 Å apart), there appears an anticorrelation between near-neighbor secondary structure elements, while at further distances (>25 Å apart), the next-near-neighbor groups are weakly correlated. This is just the type of low-frequency motion one might expect from an assembly of rigid bodies.

In Figure 5, the rms fluctuations of α-carbons is compared for each of the eight lowest frequency modes of BPTI, as is the sum of these eight modes with the sum over all 208 modes. It is clearly seen that the eight lowest modes account for most of the α-carbon displacements (70%). The next eight modes contribute an additional 12%, while the 50 lowest modes account for over 95% of the total rms fluctuations. As a result, the detailed motion of the eight lowest modes are the most interesting. Vectorial representations of these eight modes are presented in stereo diagrams in Figure 6. These modes have also been analyzed using interactive computer graphics. As already shown in Figure 5, the low-frequency modes are collective (i.e., they involve simultaneous motion of all parts of the protein). Figure 6 shows as well that they are segmental in that segments of secondary structure or loops tend to move as one piece. This was also implied in Figure 4 by the strong positive correlation between neighboring residues.

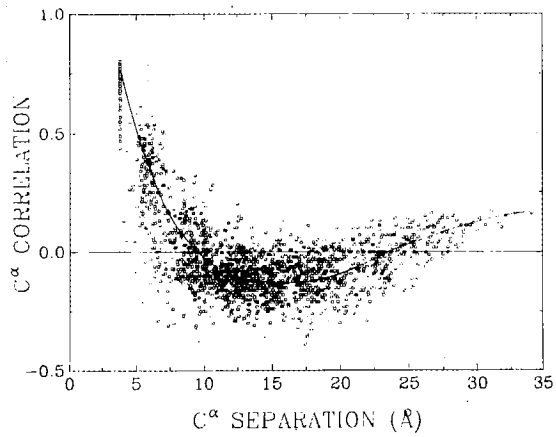
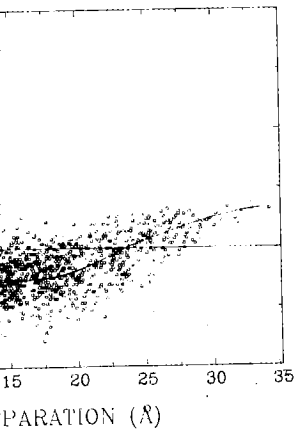



Figure 4. The $C^\alpha$ atom pair correlation is plotted as a function of the $C^\alpha$ atom separation. The curve can be interpreted as follows: Segments of secondary structure or loops of about 10 Å length tend to move together and are anticorrelated with near-neighbor segments 10 to 25 Å distant, and are weakly correlated with next-near-neighbor segments > 25 Å away.

7], and ferrocytochrome *c* [18] all yield

of α-carbon atoms plotted against the
uperimposed "eyeball" best-fit curve can
e to one another (<10 Å apart), such as
structure (e.g., α-helix or β-sheet), are
distances (~10–25 Å apart), there ap-
neighbor secondary structure elements,
art), the next-near-neighbor groups are
e of low-frequency motion one might
s.

α-carbons is compared for each of the
as is the sum of these eight modes with
seen that the eight lowest modes account
(70%). The next eight modes contribute
modes account for over 95% of the total
ed motion of the eight lowest modes are
ations of these eight modes are presented
modes have also been analyzed using
dy shown in Figure 5, the low-frequency
e simultaneous motion of all parts of the
they are segmental in that segments of
ove as one piece. This was also implied
rrelation between neighboring residues.




lotted as a function of the Cα atom separation.
gments of secondary structure or loops of about
nticorrelated with near-neighbor segments 10 to
h next-near-neighbor segments > 25 Å away.



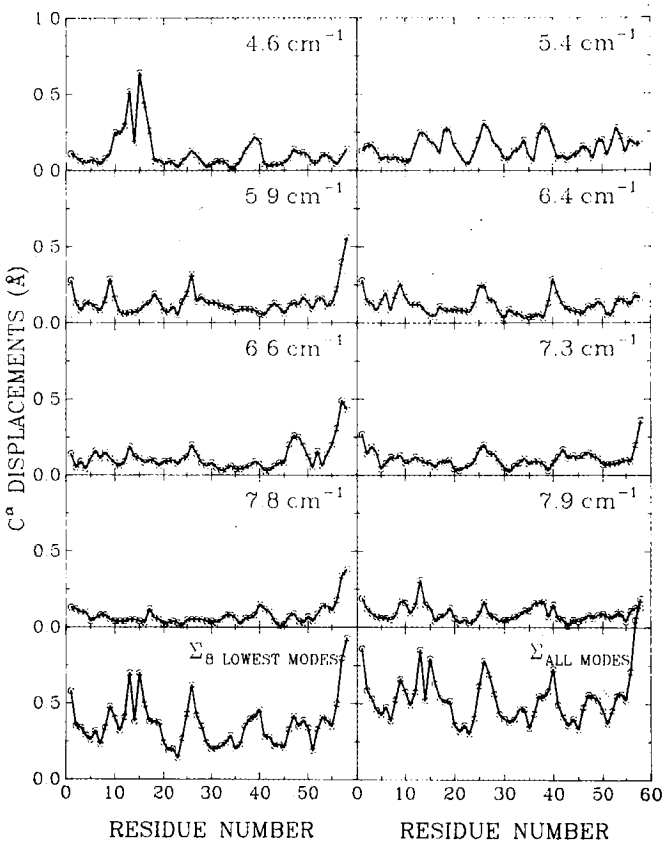
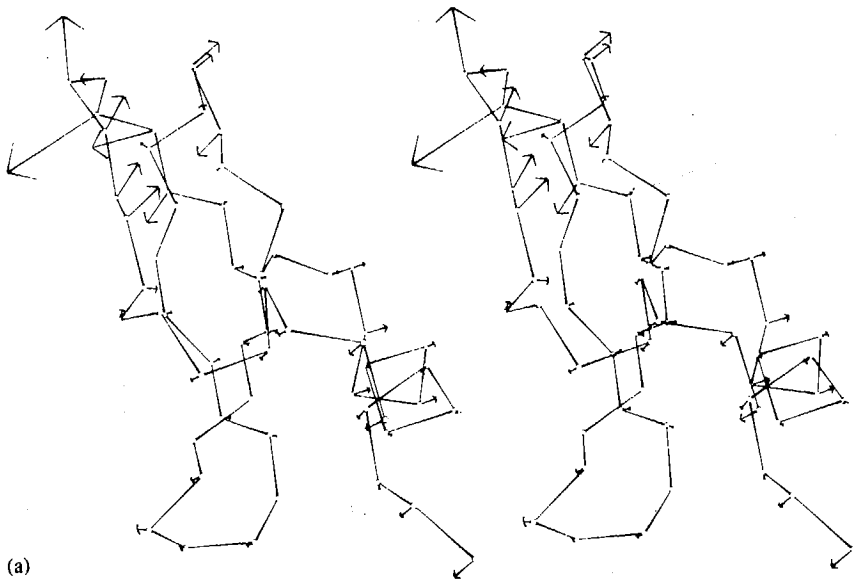



Figure 5. The Cα fluctuations in the eight lowest modes are compared. The two bottom plots show that these eight modes contribute about 70% to the total rms Cα displacements.

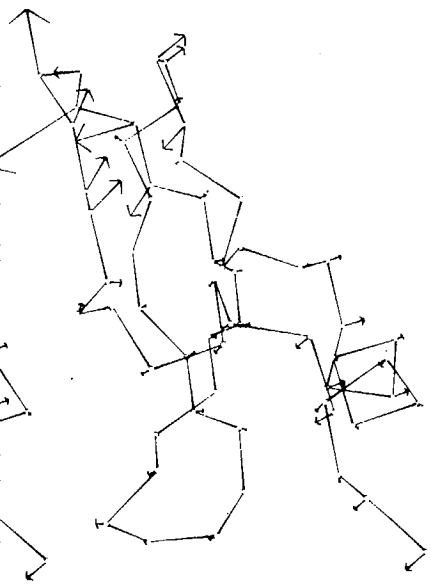
Computer graphics shows that there is little side chain or peptide unit motion within a segment.

Figure 7 reproduces the plot of the rms fluctuations versus frequency of all the 208 $\phi$, $\psi$, and $\chi$ angles as well as the contribution to various portions of the frequency spectrum of the four different types of torsion angles, i.e., those found in α-helices, β-sheets, turns, or side chains. It is seen that torsional motion in β-sheets and turns is the predominant contribution up to 45 $cm^{-1}$. From 40 to 75 $cm^{-1}$, the largest torsional motion is in turns and in side chains, while from 70 to 150 $cm^{-1}$, the largest contributions come from the α-helix and side-chain torsion angles. The band of α-helix frequencies in Figure 7, with peaks of intensity at 70 and 105 $cm^{-1}$ (Fig. 1), are somewhat higher than the 45 to 60 $cm^{-1}$ predicted for the longitudinal acoustical vibrations of a ten residue α-helix [35] but are nevertheless reasonable.



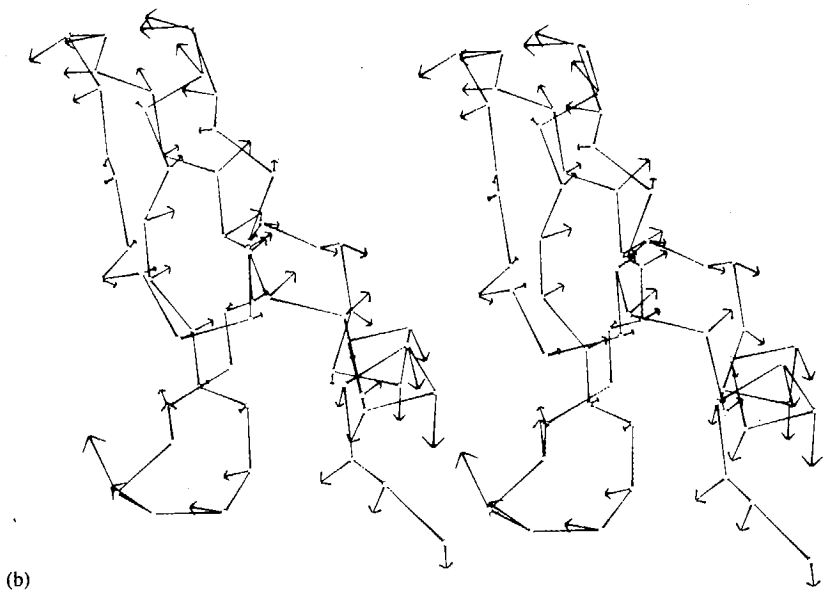
(a)

Figure 6. Stereo representation of the eight lowest normal modes of BPTI. Only the $C^{\alpha}$ backbone is shown. The length of the arrows give the rms magnitudes of the $C^{\alpha}$ fluctuations exaggerated by a factor of 10. (a) *Rotating tristar*. The largest displacements involve rotation of the tristar at the top formed by segments Gly12-Ala16 and the Cys14-Cys38 SS-bridge. Segment Pro8-Thr11 of the N-terminal arm moves up and down much like a piston connected to the tristar. (b) *Bend*. The top and bottom of the β-hairpin are moving left as the middle moves right. The entire molecule appears to bend. The motion can also be described as a gentle rocking of the bottom of the hairpin plus the C-terminal helix as a unit relative to the top half of the protein. Here, as in other modes, the displacement of segments near each other in space is positively correlated. (c) *Sliding feet*. The bottom part of the molecule, near Arg1, Lys26, and Gly56-Ala58, slides left-right as a unit; residues around Pro9 move up-down in concert. The top and back of the molecule are relatively "quiet". (d) *Opening chest*. The opposing motion of Ala40 and Tyr10 mimic the opening and closing of the front of the molecule. Similarly, as Pro9 moves down, the bottom of the hairpin loop moves up, closing the cleft between them. (e) *Skewered helix*. The C-terminal helix rotates around an axis roughly perpendicular to its axis through the helix center. Pro13 and Lys26 at the top and bottom are anticorrelated, as if the rest of the protein counter-rotates relative to the helix. (f) *Scissors*. The bottom of the hairpin loop (near Lys26) is in scissor-like motion relative to the N- plus C-terminus (Arg1, Asp3, Gly57, Ala58). There is weak opening of the protein by the opposing motion of Tyr10 and Arg42. (g) *Helix wag*. This mode is dominated by the α-helix wagging like a tail, i.e., in rotation about an axis perpendicular to the helix axis near its N-terminus, giving large amplitudes to Arg53-Ala58. The segments near Ala40-Arg42, Arg17, and Arg1-Phe4, all near the right of the molecule, also participate. (h) *Twist*. This mode gives the impression of an overall twist: while the tristar (Gly12-Lys15, Gly36-Ala40) twists in phase with the bottom of the hairpin (near Lys26), the center of the β-sheet (near Ilu19 and near Phe33) is opposed in phase. Again, there is opening by anticorrelation of Pro9-Tyr10 relative to Ala40.

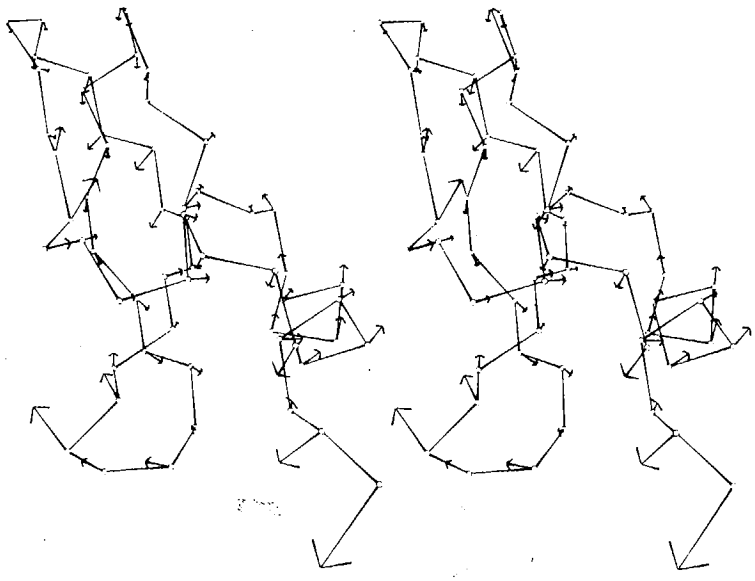


lowest normal modes of BPTI. Only the $C^{\alpha}$ give the rms magnitudes of the $C^{\alpha}$ fluctuations *star*. The largest displacements involve rotation Gly12-Ala16 and the Cys14-Cys38 SS-bridge. oves up and down much like a piston connected of the β-hairpin are moving left as the middle bend. The motion can also be described as a us the C-terminal helix as a unit relative to the odes, the displacement of segments near each *ling feet*. The bottom part of the molecule, near ight as a unit; residues around Pro9 move up- lecule are relatively "quiet". (d) *Opening chest*. mic the opening and closing of the front of the e bottom of the hairpin loop moves up, closing The C-terminal helix rotates around an axis helix center. Pro13 and Lys26 at the top and ie protein counter-rotates relative to the helix. (near Lys26) is in scissor-like motion relative 7, Ala58). There is weak opening of the protein (g) *Helix wag*. This mode is dominated by the out an axis perpendicular to the helix axis near rg53-Ala58. The segments near Ala40-Arg42, the molecule, also participate. (h) *Twist*. This while the tristar (Gly12-Lys15, Gly36-Ala40) n (near Lys26), the center of the β-sheet (near Again, there is opening by anticorrelation of



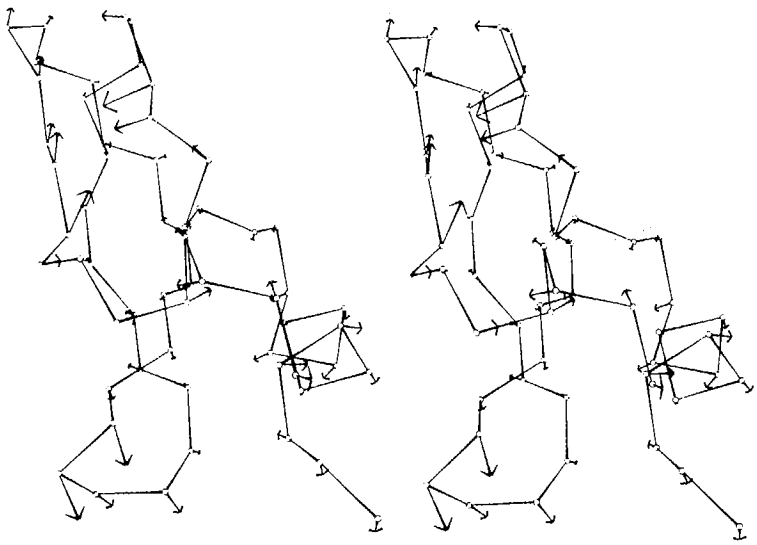


(b)

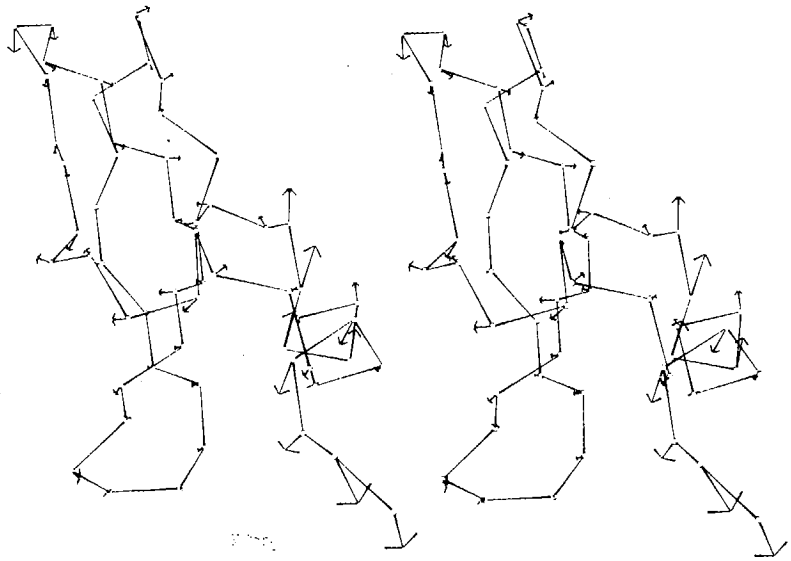


(c)

Figure 6. *(Continued from previous page.)*

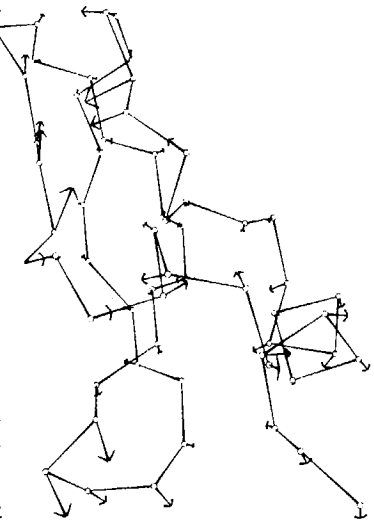


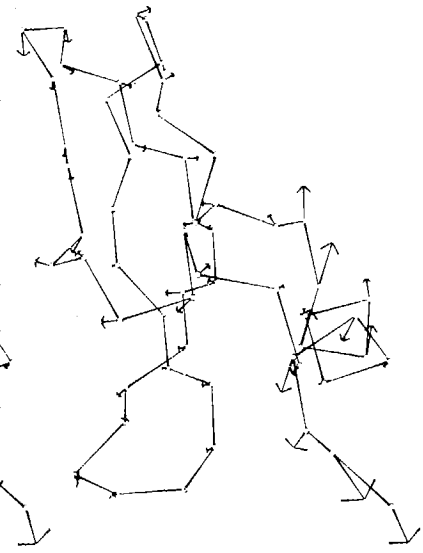
(d)



(e)

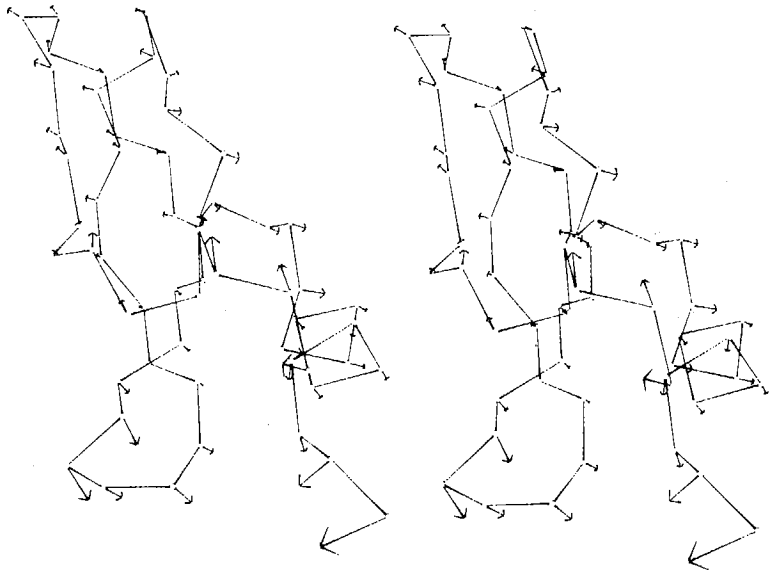
Figure 6. *(Continued from previous page.)*



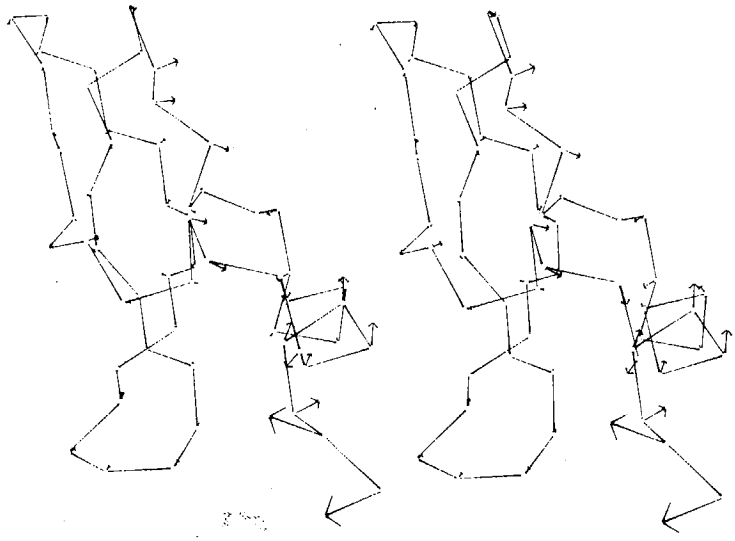


*ued from previous page.)*



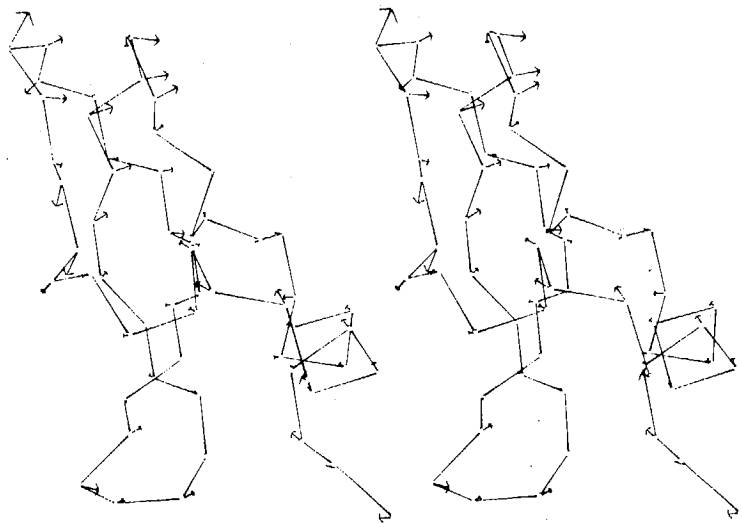


(f)



(g)

Figure 6. *(Continued from previous page.)*



(h)

Figure 6. *(Continued from previous page.)*

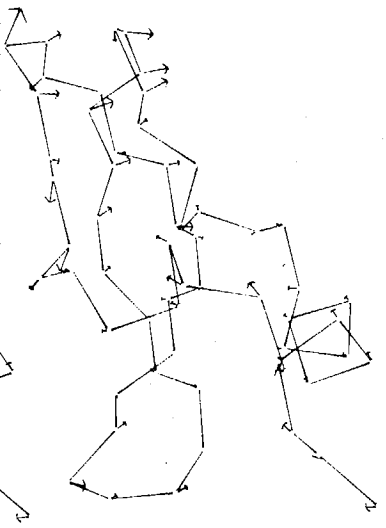
Correlations of the $\phi$ and $\psi$ angles in the $\alpha$-helix corroborate those of previous calculations [32,33]. For $\phi$, there is a weak correlation for the two nearest residues, while for $\psi$, only the next-nearest residues are weakly correlated. The $\phi_i\psi_{i+k}$ correlation is $-0.73$ for $k = -1$ and falls off for $0 \leq k < -1$. In the $\beta$-sheet, there is little correlation for $\phi_i\phi_j$ or $\psi_i\psi_j$, but $\phi_i\psi_{i-1}$ are again strongly anticorrelated ($-0.82$).

## Conclusions

A simple model has been presented which describes well the basic features of protein dynamics within the limitation of its approximations of harmonic behavior, a rigid-body potential, and the absence of solvent. The net result of these approximations is likely to be smaller amplitudes and slightly higher frequencies of vibration than would occur in nature. The main results of the calculation can be summarized as follows:

1. The key feature of normal mode analysis is decomposition of the motion into different time scales. High frequencies are associated with short wavelengths, low frequencies with long wavelengths. While considerable motion of peptide units occurs only at higher frequencies, overall motion of chain segments occurs only at low frequencies. The most significant result of this analysis, lies in the realization that one can capture the essential features of overall chain motion by concentrating on a small fraction of all possible modes. The conceptual simplification and resulting economy of computing and analysis is considerable.

2. There is convincing evidence for segmental motion, dominated by the low-frequency modes. The anticorrelated residue pairs in Figure 4 cluster in regions



*d from previous page.)*

the α-helix corroborate those of previous
a weak correlation for the two nearest
earest residues are weakly correlated. The
$-1$ and falls off for $0 \leq k < -1$. In the
$\phi_i\phi_j$ or $\psi_i\psi_j$, but $\phi_i\psi_{i-1}$ are again strongly

## clusions

d which describes well the basic features
ation of its approximations of harmonic
the absence of solvent. The net result of
maller amplitudes and slightly higher fre-
ur in nature. The main results of the cal-
s:

analysis is decomposition of the motion
uencies are associated with short wave-
avelengths. While considerable motion of
quencies, overall motion of chain segments
ost significant result of this analysis, lies
re the essential features of overall chain
ction of all possible modes. The conceptual
of computing and analysis is considerable.
segmental motion, dominated by the low-
esidue pairs in Figure 4 cluster in regions



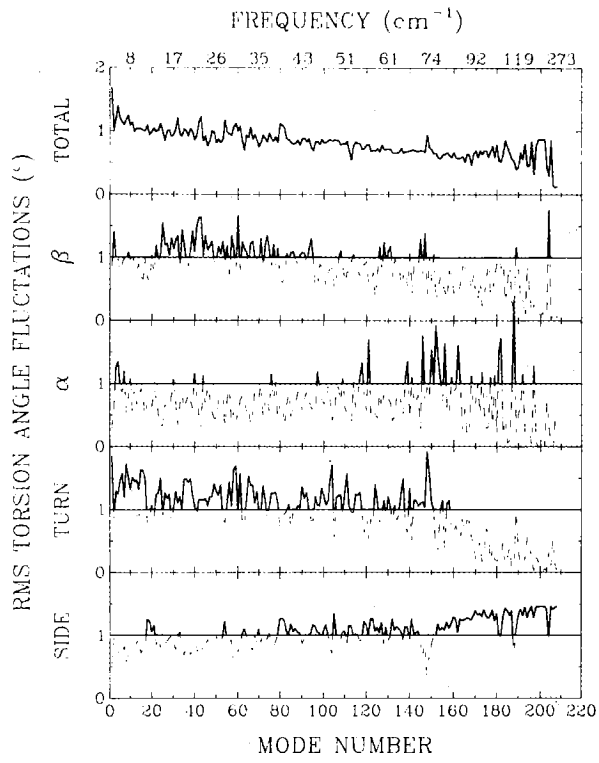



Figure 7. The rms fluctuations for the four classes of torsion angle are plotted as their ratio to the total rms fluctuations vs. frequency. Contributions from a particular torsion angle greater than the average are shown as solid lines, while contributions less than the average are shown as dotted lines.

of width 5–12 residues. The distance between large positive and large negative correlation is at least 5 residues. In the low-frequency modes (Fig. 6), the vectors change direction with a spatial frequency of about 10 Å. A color graphics film of the low-frequency modes gives the distinct impression of segmental coherence (see description above). While the size of motional segments may change somewhat for larger proteins, and while the details of the lowest modes certainly will change for different proteins, segmental motion is expected to be a characteristic feature of all low frequency vibrations of globular proteins.

3. As the low-frequency modes are mainly responsible for the rms displacement of $C^\alpha$ atoms, it is possible that the initial stages of unfolding involve motion that can be described as a superposition of the lowest few modes. Any significant unfolding must involve rearrangement of solvent molecules. Are the slowest normal mode vibrations slow enough to facilitate solvent rearrangement or are they so rapid that the solvent sees only the average position of protein groups? In the present approximation, the slowest modes have a period of the order of 10 ps. From the known diffusion constants of water, one can estimate a rotational

diffusion time of $10^{-13}$ and a translational diffusion of about 1.5 Å per 10 ps; thus, significant water motion is possible during one cycle of a collective normal mode of BPTI at room temperature, and certainly also at the higher unfolding temperature. Thus, it may be reasonable to try to model unfolding as a process involving mainly slow collective modes.

## Bibliography


[1] M. Levitt, Ann. Rev. Biophys. Bioeng. **11,** 251 (1982).
[2] P. G. Debrunner and H. Frauenfelder, Ann. Rev. Phys. Chem. **33,** 283 (1982).
[3] M. Karplus and J. A. McCammon, CRC Crit. Rev. Biochem. **9,** 293 (1981).
[4] J. A. McCammon and M. Karplus, Ann. Rev. Phys. Chem. **31,** 29 (1980).
[5] A. Cooper, Sci. Prog. **66,** 473 (1980).
[6] F. R. N. Gurd and T. M. Rothgeb, Adv. Protein Chem. **33,** 74 (1979).
[7] W. L. Peticolas, Methods Enzymol. **61,** 425 (1979).
[8] W. N. Lipscomb, Acc. Chem. Res. **3,** 81 (1970).
[9] M. F. Perutz, Nature (London) **228,** 726 (1970).
[10] R. Huber, J. Deisenhofer, P. M. Colman, M. Matsushima, and W. Palm, Nature (London) **264,** 415 (1976).
[11] M. F. Perutz, Ann. Rev. Biochem. **48,** 327 (1979).
[12] M. Ataka and S. Tanaka, Biopolymers **18,** 507 (1979).
[13] K. G. Brown, S. C. Erfurth, E. W. Small, and W. L. Peticolas, Proc. Natl. Acad. Sci. U.S.A. **69,** 1467 (1972).
[14] L. Genzel, R. Keilmann, T. P. Martin, G. Winterling, Y. Yacoby, H. Fröhlich, and M. Makinen, Biopolymers **15,** 219 (1976).
[15] P. C. Painter, L. E. Mosher, and C. Rhoads, Biopolymers **21,** 1469 (1982).
[16] H. Frauenfelder, G. A. Petsko, and D. Tsernoglou, Nature (London) **280,** 558 (1979).
[17] P. J. Artymiuk, C. C. F. Blake, D. E. P. Grace, S. J. Oatley, D. C. Phillips, and M. J. E. Sternberg, Nature (London) **280,** 563 (1979).
[18] S. H. Northrup, M. R. Pear, J. A. McCammon, M. Karplus, and T. Takano, Nature (London) **287,** 659 (1980).
[19] I. Munro, I. Pecht, and L. Stryer, Proc. Natl. Acad. Sci. U.S.A. **76,** 56 (1979).
[20] J. R. Lakowicz and G. Weber, Biochemistry **12,** 4171 (1973).
[21] C. K. Woodward and B. D. Hilton, Ann. Rev. Biophys. Bioeng. **8,** 99 (1979).
[22] R. Richarz, P. Sehr, G. Wagner, and K. Wüthrich, J. Mol. Biol. **130,** 19 (1979).
[23] B. D. Hilton and C. K. Woodward, Biochemistry **18,** 5834 (1979).
[24] I. D. Campbell, C. M. Dobson, and R. J. P. Williams, Adv. Chem. Phys. **39,** 55 (1978).
[25] A. A. Ribeiro, R. King, C. Restivo, and O. Jardetzky, J. Am. Chem. Soc. **102,** 4040 (1980).
[26] J. S. Valentine, R. P. Sheridan, L. C. Allen, and P. C. Kahn, Proc. Natl. Acad. Sci. U.S.A. **76,** 1009 (1979).
[27] A. Warshel and M. Levitt, J. Mol. Biol. **103,** 227 (1976).
[28] Y. Suezaki and N. Gō, Int. J. Pept. Protein Res. **7,** 333 (1975).
[29] N. Gō, Biopolymers **17,** 1373 (1978).
[30] Y. Ueda and N. Gō, Int. J. Pept. Protein Res. **8,** 551 (1976).
[31] T. Noguti and N. Gō, Nature (London) **296,** 776 (1982).
[32] R. M. Levy and M. Karplus, Biopolymers **18,** 2465 (1979).
[33] M. Gō and N. Gō, Biopolymers **15,** 1119 (1976).
[34] Y. Suezaki and N. Gō, Biopolymers **15,** 2137 (1976).
[35] W. C. Peticolas, Biopolymers **18,** 747 (1979).
[36] J. A. McCammon, B. R. Gelin, and M. Karplus, Nature (London) **267,** 585 (1977).
[37] M. Karplus and J. A. McCammon, Nature (London) **277,** 578 (1979).
[38] M. Levitt, J. Mol. Biol. **168,** 595, 621 (1983).

nal diffusion of about 1.5 Å per 10 ps;
e during one cycle of a collective normal
nd certainly also at the higher unfolding
le to try to model unfolding as a process
s.

## graphy


1, 251 (1982).
n. Rev. Phys. Chem. **33,** 283 (1982).
Crit. Rev. Biochem. **9,** 293 (1981).
Rev. Phys. Chem. **31,** 29 (1980).

Protein Chem. **33,** 74 (1979).
25 (1979).
(1970).
1970).
, M. Matsushima, and W. Palm, Nature (London)

27 (1979).
3, 507 (1979).
and W. L. Peticolas, Proc. Natl. Acad. Sci. U.S.A.

G. Winterling, Y. Yacoby, H. Fröhlich, and M.

ads, Biopolymers **21,** 1469 (1982).
sernoglou, Nature (London) **280,** 558 (1979).
. Grace, S. J. Oatley, D. C. Phillips, and M. J. E.
79).
amon, M. Karplus, and T. Takano, Nature (London)

Natl. Acad. Sci. U.S.A. **76,** 56 (1979).
stry **12,** 4171 (1973).
. Rev. Biophys. Bioeng. **8,** 99 (1979).
Wüthrich, J. Mol. Biol. **130,** 19 (1979).
chemistry **18,** 5834 (1979).
. P. Williams, Adv. Chem. Phys. **39,** 55 (1978).
O. Jardetzky, J. Am. Chem. Soc. **102,** 4040 (1980).
llen, and P. C. Kahn, Proc. Natl. Acad. Sci. U.S.A.

**103,** 227 (1976).
tein Res. **7,** 333 (1975).

n Res. **8,** 551 (1976).
**296,** 776 (1982).
rs **18,** 2465 (1979).
9 (1976).
2137 (1976).
1979).
Karplus, Nature (London) **267,** 585 (1977).
ure (London) **277,** 578 (1979).
(1983).





[39] S. Lifson and A. Warshel, J. Chem. Phys. **49,** 5116 (1968).
[40] E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw-Hill, New York, 1955).
[41] A. W. Burgess and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **72,** 1221 (1975).
[42] B. Robson, P. S. Stern, I. H. Hillier, D. J. Osguthorpe, and A. T. Hagler, J. Chim. Phys. **76,** 831 (1979).
[43] P. S. Stern, M. Chorev, M. Goodman, and A. T. Hagler, Biopolymers **22,** 1885 (1983).
[44] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Massachusetts, 1950).
[45] M. Levitt, J. Mol. Biol., to appear.
[46] R. Huber, D. Kukla, A. Ruhlman, O. Epp, and H. Formanek, Naturwissenschaften **57,** 389 (1970).
[47] J. Deisenhofer and W. Steigemann, Acta Crystallogr. **B31,** 238 (1975).
[48] R. Fletcher, U.K. Atomic Energy Authority, Harwell, Report AERE R7125, 1972.
[49] M. Levitt, Ph.D. thesis, University of Cambridge, 1972.
[50] C. Sander and P. S. Stern, *Workshop du CECAM: Analyse des Structures des Proteines,* Rapport d'Activité Scientifique du CECAM, Université Paris-Sud, Orsay, 1979, pp. 121–127.
[51] H. Katz, R. Walter, and R. L. Somorjai, Computers and Chemistry **3,** 25 (1979).
[52] W. Kabsch, Acta Crystallogr. **A32,** 922 (1976).
[53] A. D. McLachlan, J. Mol. Biol. **128,** 49 (1979).
[54] J. H. Wilkinson and C. Reinsch, *Handbook for Automatic Computation Vol. II, Linear Algebra* (Springer-Verlag, Berlin, 1971).
[55] C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Mécanique Quantique,* Vol. 1 (Hermann, Paris, 1973), p. 634.
[56] A. T. Hagler, E. Huler, and S. Lifson, J. Am. Chem. Soc. **96,** 5319 (1974).
[57] P. S. Stern, M. Chorev, M. Goodman, and A. T. Hagler, Biopolymers **22,** 1901 (1983).


Received March 15, 1983