Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 6571–6575, November 1983
Biophysics

Harmonic dynamics of proteins: Normal modes and fluctuations in bovine pancreatic trypsin inhibitor

(vibrational analysis/protein dynamics/temperature factors/thermodynamic properties/frequency spectrum)

BERNARD BROOKS AND MARTIN KARPLUS

Department of Chemistry, Harvard University, Cambridge, MA 02138

Contributed by Martin Karplus, July 5, 1983

ABSTRACT **A normal mode analysis making use of an empirical potential function including local and nonlocal (nonbonded) interactions is performed for the bovine pancreatic trypsin inhibitor in the full conformational space of the molecule (1,740 degrees of freedom); that is, all bond lengths and angles, as well as dihedral angles, are included for the 580-atom system consisting of all heavy atoms and polar hydrogens. The heavy-atom frequency spectrum shows a dense distribution between 3 and 1,800 cm^{-1}, with 350 modes below 216 cm^{-1}. Most of the low-frequency modes, of which many have significant anharmonic character, are found to be delocalized over the protein. The root-mean-square amplitudes of the atomic fluctuations are calculated at 300 K from the normal modes and compared with those obtained from a solution molecular dynamics simulation based on the same potential function; very good agreement is obtained for the variation in the main-chain fluctuations as a function of residue number, though larger differences occur for the side chains. The fluctuations are generally, though not always, dominated by frequencies below 30 cm^{-1}, in accord with the results of the dynamics simulation. The vibrational contributions to the thermodynamic properties of the protein are calculated as a function of temperature; the effects of perturbations on the spectrum, suggested for ligand or substrate binding, are examined. The analysis demonstrates that, in spite of the anharmonic contributions to the potential, a normal mode description can provide useful results concerning the internal motions of proteins.**

The nature of the internal motions of proteins is a subject of considerable interest, particularly because some of these motions are known to play an important role in protein function (1, 2). Various theoretical and experimental methods are now being employed to determine both the magnitudes and the time scales of the internal motions (1–4). Molecular dynamics, in particular, has been shown to be a powerful approach for the study of fluctuations on the subnanosecond time scale (2, 3). This is true because it is based on integrating the equations of motion for the individual atoms with the exact forces corresponding to a given potential function. Simulations for proteins have shown (2, 5) that the atomic fluctuations can be separated into local oscillations superposed on motions with a more collective character. The former have a subpicosecond time scale; the latter vary from 1 to 10 ps or longer, corresponding to frequencies in the range 30 to 3 cm^{-1}. It is these collective motions that introduce the variation in the magnitude of the fluctuation that characterize different parts of a protein and are likely to be of primary importance in biological function.

To examine the collective character of the fluctuations in more detail, it is useful to apply an alternative approach, harmonic dynamics, for analyzing the internal motions of proteins. In harmonic dynamics (6) it is assumed that the fluctuations of atoms are sufficiently small that the potential energy can be approximated as a sum of terms that are quadratic in the displacements. Although this is known to be an approximation (i.e., molecular dynamics simulations have shown that the fluctuations have anharmonic contributions) (7–9), a harmonic analysis can, nevertheless, provide insights concerning the motional behavior. The frequencies and forms of the normal modes are determined by diagonalizing the mass-corrected force constant matrix. Once the frequencies and modes are known, the magnitudes, times scales, and correlations of the atomic fluctuations can be calculated at any temperature by summing over the modes. Also, the harmonic model results are of direct interest for comparison with a variety of spectroscopic measurements (infrared, Raman, neutron scattering) and permit evaluation of quantum effects on the fluctuations and thermodynamic properties.

In this paper, we develop the harmonic model for proteins and apply it to the bovine pancreatic trypsin inhibitor (BPTI). This protein was chosen for study because of the wealth of the available experimental and theoretical information. We use the same form of potential as was employed in the molecular dynamics simulations (9, 10) and include all of the degrees of freedom of the molecule; in the extended atom model with explicit treatment of the polar hydrogens, there are 1,734 vibrational frequencies, which are found to span a range from ≈3 to ≈3,000 cm^{-1}. Because the primary concern is with the heavy atom fluctuations, we limited the determination of the normal modes to the set of 900 with the lowest frequencies (less than 732 cm^{-1}). Special methods were developed for determining the vibrational frequencies and normal modes of such a large system. The approach developed here is general and has been applied to other large molecules such as DNA (11). A preliminary report, including a film of some BPTI modes, has been presented previously (12).

Two recent papers have been concerned with simplified treatments of proteins in the harmonic model; one of these (13) employed an approximate representation of the glucagon monomer and used a local force field (i.e., nonlocal interaction terms were neglected), whereas the other (14) used a full potential analogous to the one employed here but diagonalized the force constant matrix without mass-weighting and restricted the calculation to the space of dihedral angles.

METHOD

BPTI was treated as a system of 580 atoms; that is, 458 heavy atoms, including 4 interior water oxygens, and 122 polar hydrogen atoms capable of forming hydrogen bonds; the remaining hydrogens were treated as part of the heavy atom to which

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: BPTI, bovine pancreatic trypsin inhibitor.

they are attached (extended atom model). The potential energy function used is essentially that included in a general program (CHARMM) (10) for energy minimization and molecular dynamics; it consists of local terms associated with bond lengths, bond angles, dihedral angles, and improper dihedral angles and nonlocal terms associated with van der Waals, electrostatic, and hydrogen bonding interactions; a distance-dependent dielectric constant was used for the electrostatic terms, which were summed without long-range truncation.

To obtain a structure for the normal mode calculation, it was necessary to refine the x-ray crystal structure (15). Although it is usual to carry out normal mode calculations for the minimal energy conformation, for proteins energy minimization can give rise to problems; in particular, the energy-minimized structure is expected to approximate the contracted zero-temperature structure, whose dynamic properties may differ significantly from the room-temperature form (16). This is one of the reasons that the molecular dynamics approach, in which the structural effects of kinetic energy are included (i.e., static potentials are replaced by potentials of mean force), is appropriate for studying the fluctuations and other properties of bulk systems. To obtain the quadratic force field expected for the protein with its room-temperature structure, constraints were introduced that led to a minimal energy structure close to the crystal structure; a mass-weighted atomic harmonic potential was applied to each atom because it does not directly affect the normal modes. Successive sets of minimizations [500 adapted basis Newton–Raphson steps each (10)] were performed with the constraint reduced and the reference structure updated for each set; the final constraint constant was 0.05 kcal/mol·Å^2 per atomic mass (1 kcal = 4.18 kJ) and the overall root-mean-square (rms) deviation from the x-ray structure was 0.62 Å. Given the resulting structure, the normal mode and frequency calculations were made without constraints.

Because of the large dimension of the normal mode problem for BPTI (1,740 modes), special methods had to be developed for constructing and manipulating the mass-weighted second-derivative matrix and for using it to determine the vibrational frequencies and normal modes. Unlike other large matrix problems, where simplifications can be introduced because the diagonal elements are dominant, the force constant matrix for a globular protein has large off-diagonal elements due to the presence of nonlocal interactions in the potential function. Details of the procedure, including the need to limit computer storage and time requirements and minimize error accumulation, will be given in a separate report. All of the calculations were performed with mass-corrected Cartesian coordinates; the use of internal coordinates can introduce difficulties for such large systems with closed loops and nonlocal interactions.

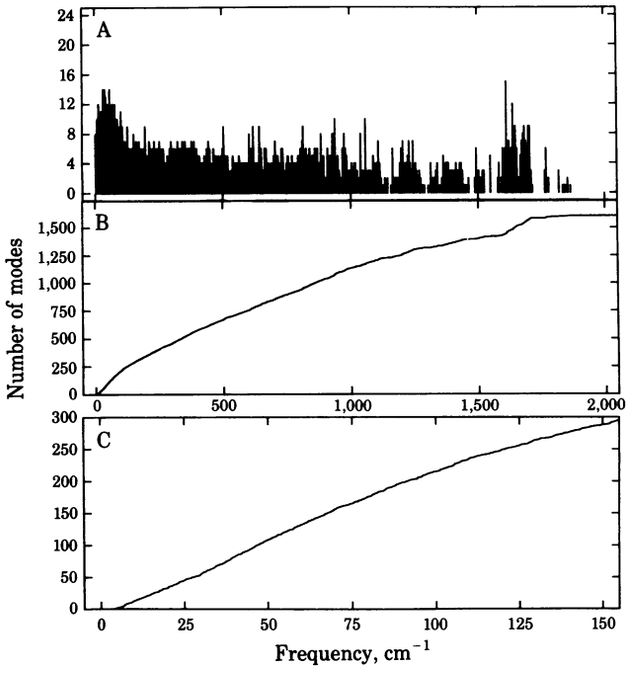

FIG. 1. Normal mode frequencies. (*A*) Histogram of the number of normal modes per 5-cm^{-1} interval (hydrogen stretches are not shown); (*B*) number of modes below a given frequency; (*C*) expanded version of *B* in the low-frequency region.

RESULTS

The histogram in Fig. 1*A* shows the calculated normal mode spectrum of BPTI, including all degrees of freedom of the 580-atom system; the figure shows all the frequencies up to 2,000 cm^{-1} (hydrogen stretching frequencies are not shown); Fig. 1*B* gives the cumulative distribution for the number of modes below a given frequency, and Fig. 1*C* shows an expanded cumulative distribution for the lowest 300 modes of primary interest. There is an essentially continuous, though not completely uniform, distribution of frequencies between 3.1 and 1,200 cm^{-1}. Between 1,200 cm^{-1} and 1,800 cm^{-1}, the frequencies tend to come in groups, many of which are dominated by bond-stretching vibrations. There are 20 modes between 3.1 and 13 cm^{-1} and there is a peak in the frequency distribution near 50 cm^{-1}. Because the structure used was not an absolute minimum (see above), 7 negative modes were found; energy searches along these modes, which are all local in character, indicated that their

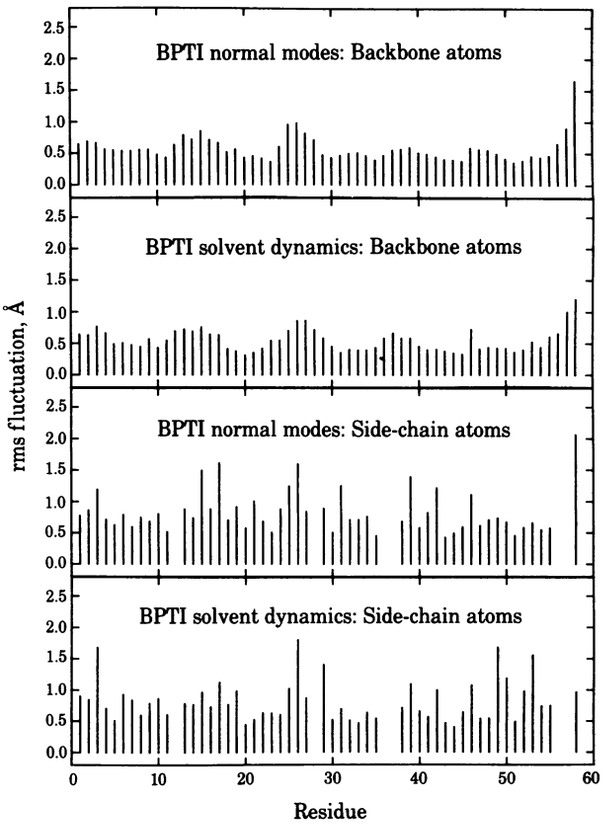

FIG. 2. rms atomic fluctuations of BPTI at 300 K averaged over each residue from normal mode and molecular dynamics; separate plots are given for main-chain (N, C$^\alpha$, C) and side-chain heavy atoms.

Table 1. Atom-type averaged rms fluctuations

Type	rms fluctuations, Å		
	Normal modes	Adjusted modes*	Dynamics
All atoms	0.776	0.555	0.714
Main chain	0.598	0.424	0.582
Side chain	0.905	0.643	0.846
C^{α}	0.611	0.436	0.536
C^{β}	0.712	0.517	0.662
γ position	0.736	0.540	0.768
δ position	0.841	0.626	0.912
ε position	0.996	0.837	0.970

*This column uses normal mode frequencies based on an energy search (see text).

correct frequencies are in the range 20 to 40 cm^{-1}.

The rms atom fluctuations were calculated from the normal modes by evaluation of the classical expression (6, 7)

$$\langle \Delta r_k^2 \rangle = k_B T \sum_i \frac{|\vec{a}_{ik}|^2}{\omega_i^2},$$

in which $\vec{a}_{ik}$ is the vector of the projections of the ith normal mode with frequency ω_i on the Cartesian components of the displacement vector for the kth atom, k_B is the Boltzmann constant, and T is the absolute temperature; quantum corrections are negligible above 50 K (7). Fig. 2 shows the normal mode rms fluctuations calculated at 300 K and compares them with the results of a molecular dynamics simulation of BPTI in a van der Waals solvent (9); this simulation was used because its average structure is closest to that employed for the normal mode analysis. The results for main-chain and side-chain averages as a function of residue number are given. For the main-chain fluctuations, the molecular dynamics and normal mode values are very similar; for the side chains, there is also a correspondence, though the differences are more pronounced. The main chain values show that the COOH terminus has large fluctuations, as does the loop region at the bottom of the molecule (residues 25–29) and the binding site in the neighborhood of residues 14 and 38 at the top of the molecule. By contrast the β-sheet residues (18–24, 29–35) show smaller fluctuations; the α-helices (3–7, 47–56) are intermediate.

The origin of the differences between the molecular dynamics and normal mode results is likely to have contributions from anharmonic and solvent effects and from the difference between the average dynamics structure and that used for the normal mode analysis. The main-chain atoms apparently experience a potential of mean force that is closer to the harmonic potential than do the side chains; because the dynamics simulation was done in a van der Waals solvent, the exterior side chains are expected to be most perturbed. To investigate the nature of the potential in the neighborhood of the structure used for the harmonic model, we have done an energy search along the low-frequency normal-mode displacements. The energy dependence showed some anharmonic contributions. Fitting the resulting energies to a parabola generally led to an increase in the effective frequencies, which reduces the rms fluctuations (see Table 1); the relative values of the atomic fluctuations are essentially unchanged.

Analysis of the time scale as well as the magnitude of the fluctuations has been made for the molecular dynamics results.

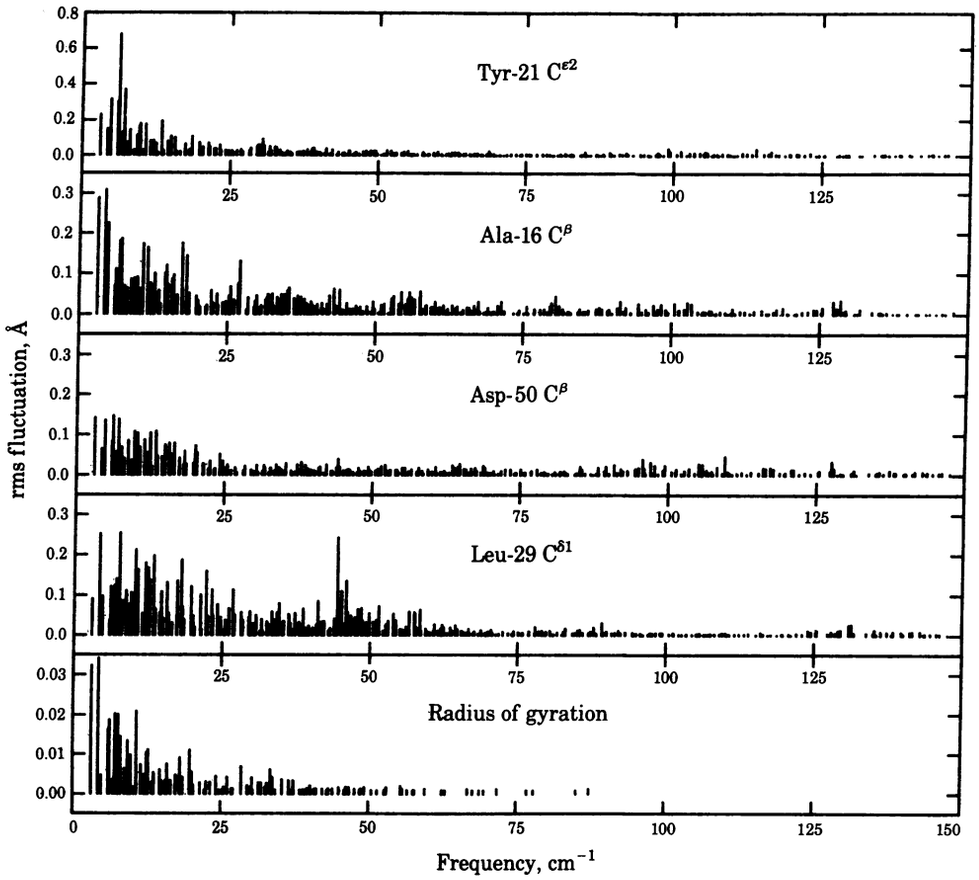

FIG. 3. Contribution of normal modes to rms fluctuations as a function of frequency; selected atoms and the radius of gyration are included.

From the calculated time series and correlation functions (5), it has been found that the atomic motions contributing to the rms displacements generally have a small local high-frequency component ($\approx$0.2 ps or $\approx$150 cm^{-1}) on which are superposed motions of a more collective character with time scales ranging from 1 to 10 ps or 30 to 3 cm^{-1}. The present normal mode study essentially confirms the dynamics results. Fig. 3 shows the contributions of the different normal modes to the displacements of some of the atoms whose motions were analyzed in the molecular dynamics simulations (5); also included is the fluctuation of the radius of gyration for the molecule. In most cases, the dominant contributions come from low-frequency modes in the range 3 to 50 cm^{-1}, although nonnegligible contributions come from higher frequencies up to 130 cm^{-1}. It is evident that for certain atoms (e.g., $C^{\epsilon 2}$ of Tyr-21), only a very small number of modes are important, whereas for other atoms (e.g., Ala-16 C^{β}, Asp-50 C^{β}) a range of frequencies is involved; for Leu-29 $C^{\delta 1}$, a mode at 44.5 cm^{-1} makes a very large contribution.

It is of considerable interest to examine the form of the normal modes themselves. This is of particular importance for the evaluation of the correlation between the motions of different atoms and different groups of atoms. Analysis of the dynamics results (5) has indicated that the larger scale motions have a collective character that may involve a few neighboring atoms, a residue, or groups of many atoms in a given region of a protein. We show in Fig. 4 the distribution of the displacements over residues of some of the low-frequency modes. Also included is one of the translation modes, which clearly demonstrates the purity of this mode, a pictural verification of the accuracy of the normal mode determination procedure. Most of the 300 lowest modes are highly delocalized; they are generally distributed over the entire molecule. A striking exception is one of the "negative" modes that is localized in the loop region. The lowest real mode (at 3.1 cm^{-1}) mirrors the overall rms fluctuation (see Fig. 2). Other modes shown, although they are also delocalized, are distributed somewhat differently over the various portions of the molecule. In considering the character of the individual modes, it must be recognized that because of the close spacing, relatively small effects, such as solvent damping or external perturbations (e.g., ligand binding), can lead to significant mode mixing. This may be of biological interest. It also suggests that, rather than individual mode properties, those that involve averages over a range of modes with similar frequencies are likely to be most significant and least sensitive to anharmonic corrections.

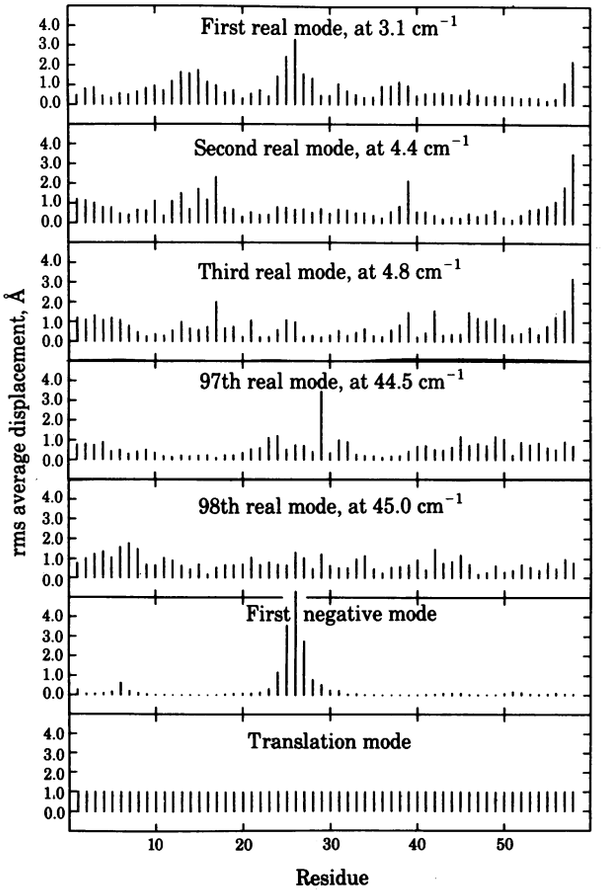

FIG. 4. Normal mode distribution: rms average displacement of atoms within a residue for a 1-Å rms displacement along selected modes.

Given the results of the present analysis (frequencies and normal modes) it is possible to calculate a wide range of experimental properties, always subject to the caveat that a harmonic model is being used. Spectral results such as infrared and Raman frequencies follow directly, though the broadening due to solvent damping should be included for experimental comparisons; a point-charge model can be used for the infrared intensities. Also, it is straightforward to determine the normal mode contributions to motional effects involved in nuclear magnetic resonance (17, 18), fluorescence depolarization (19, 20), and hydrogen exchange (21), all of which have been examined by molecular dynamics simulations. Here we report the results of the vibrational contribution to the thermodynamic properties. It is essential to do quantum mechanical calculations to obtain absolute values although, as has been pointed out in a quasi-harmonic formulation of molecular dynamics (22), classical values may be adequate for room temperature entropy differences between conformations. In Fig. 5 we show the vibrational contribution to the enthalpy, entropy, free energy, and heat capacity as a function of temperature; values up to 2,000 K are included only to indicate the approach to the asymptotic classical limits in the functional behavior; corrections for neglect of the nonpolar hydrogens have not been made. Of particular interest is the low-temperature heat capacity (see *Inset*), which is sensitive to the density of states in the frequency range of interest. Although no measurements on BPTI exist, heat capacity measurements have been in the range of 1 to 20 K for polyglycine (23) and between 10 and 310 K for anhydrous in-

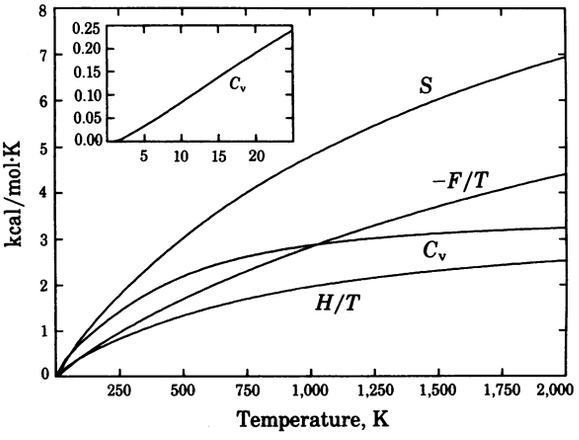

FIG. 5. Vibrational contributions to thermodynamic functions calculated from the normal mode spectrum as a function of temperature; listed are the heat capacity (C_v), entropy (S), enthalpy (H), and free energy (F). (*Inset*) C_v at low temperatures shown on an expanded scale.

sulin and α-chymotrypsin (24). Room-temperature data for a number of proteins in solution (25) yield C_p equal to 0.32 ± 0.2 cal/K·g; the calculated vibrational C_v is 0.25 cal/K·g. The calculated value of dC_v/dT at 300 K is 0.0011 cal/K²·g, whereas the measured values range from 0.001 to 0.002 cal/k²·g.

It has been suggested (26) that binding of ligands and substrates can have a significant effect on the thermodynamic properties by perturbing the low-frequency vibrations of the protein; e.g., for BPTI some change might be expected on binding to trypsin. As a model for this effect, we compare the results obtained for the directly calculated and adjusted modes, the latter corresponding to the "perturbed" system with higher frequencies. At 100 K, the vibrational free energy changes from −41.5 to −37.7 kcal/mol in the presence of the perturbation; at 300 K, the values are −336.4 and −325.1 kcal/mol, respectively. At all temperatures the vibrational enthalpy increases while the entropy decreases, leading to a significant destabilizing effect on the system. The change in enthalpy contrasts with that assumed in previous discussions (26), due to the fact that we have included the zero-point contribution. These considerations can be applied to systems such as hexokinase, for which inelastic neutron scattering has indicated that there is a change in the low-frequency spectrum on ligand binding (27).

CONCLUSIONS

This paper provides a complete normal mode analysis for a protein, BPTI. All degrees of freedom of the main-chain and side-chain atoms (bond lengths, bond angles, and dihedral angles) were included in the calculation. The present study is of special interest because it uses the same form of potential function as has been employed in a series of energy minimization and molecular dynamics studies of this molecule.

The calculated normal mode frequencies range between 3.1 and 3,200 cm^{-1}, with an almost continuous distribution below 1,200 cm^{-1}; a peak in the density of states occurs near 50 cm^{-1}. Most of the low-frequency modes are highly delocalized, suggesting that correlated fluctuations occur in the protein; this is in accord with molecular dynamics results (ref. 5 and unpublished calculations). The rms atomic fluctuations calculated from the normal modes behave similarly to those from a molecular dynamics simulation; particularly for the main chain, the variation of the fluctuations as a function of residue number is nearly identical. Further, the frequency distribution of the modes that make the dominant contributions to the fluctuations (3 to 30 cm^{-1}) is in accord with the estimates from molecular dynamics. This agreement provides evidence for the utility of a normal mode treatment for the analysis of the internal motions of proteins.

The dense distribution of modes in the low-frequency region (150 modes in the range 3 to 60 cm^{-1}) and approximately 350 modes below 200 cm^{-1}, which corresponds to k_BT at room temperature, is of considerable interest. All of these will be populated at room temperature and can serve as energy sources for structural change and enzymatic activity. It is just these low-frequency modes that tend to be highly delocalized. As already suggested from molecular dynamics results (5), this makes them candidates for the transmission of information from one part of the molecule to another, due to binding or solvent perturbations or even amino acid substitutions via natural or artificially induced mutations.

Anharmonic effects, which molecular dynamics has shown to be important in protein motions, are neglected in the normal mode calculation. In particular, a molecular dynamics simulation samples many minima in the multidimensional configuration space, whereas the harmonic model is restricted to a single minimum. The present results suggest that for certain motional properties, especially those involving averages over many modes, the harmonic model is a useful first approximation; this is certainly correct at low temperatures and in BPTI it appears true even at room temperature. Apparently, the effective potentials seen by many of the atoms or groups of atoms are sufficiently similar for the different minima that the relative values of the fluctuations are preserved. Adjustment of the force constants to account for some anharmonic contributions can extend the range of the normal mode treatment ["quasi-harmonic model" (22)]. Phenomena, such as ring flips or larger rearrangements involving barriers, are clearly outside the harmonic realm though even here the normal mode model may provide some insights. Whether it will be fruitful to approximate protein motions as involving fluctuations in multidimensional wells on a short time scale with transitions to other wells on a longer time scale awaits the results of future studies.

We thank U. Pezzeca and H. Yu for assistance with the calculations; A. Brünger for program development; and C. Brooks III, R. Bruccoleri, S. Cusack, J. Smith, and S. Swaminathan for helpful discussions. This work has been supported in part by grants from the National Science Foundation and the National Institutes of Health. We are very grateful to the Cardiac Computer Center (Massachusetts General Hospital) and its director, J. Newell, for providing the computer time that made these calculations possible.

1. Williams, R. J. P. (1979) *Biol. Rev.* **54**, 389–437.
2. Karplus, M. & McCammon, J. A. (1983) *Annu. Rev. Biochem.* **53**, 263–300.
3. Karplus, M. & McCammon, J. A. (1981) *Crit. Rev. Biochem.* **9**, 293–349.
4. Gurd, F. R. N. & Rothgeb, T. M. (1979) *Adv. Protein Chem.* **33**, 73–165.
5. Swaminathan, S., Ichiye, T., van Gunsteren, W. & Karplus, M. (1982) *Biochemistry* **21**, 5230–5241.
6. Levy, R. M. & Karplus, M. (1979) *Biopolymers* **18**, 2465–2495.
7. Levy, R. M., Perahia, D. & Karplus, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1346–1350.
8. Mao, B., Pear, M. R., McCammon, J. A. & Northrup, S. H. (1982) *Biopolymers* **21**, 1979–1990.
9. Van Gunsteren, W. F. & Karplus, M. (1982) *Biochemistry* **21**, 2259–2274.
10. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comput. Chem.* **4**, 187–217.
11. Tidor, B., Irikura, K., Brooks, B. & Karplus, M. (1983) *J. Biomol. Struct. Dyn.* **1**, in press.
12. Karplus, M. (1982) in *Structure, Dynamics, Interactions and Evolution of Biological Macromolecules*, ed. Hélène, C. (Reidel, Dordrecht, The Netherlands), pp. 253–269.
13. Tasumi, M., Takenchi, H., Ataka, S., Dwidedi, A. M. & Krimm, S. (1982) *Biopolymers* **21**, 711–714.
14. Noguti, T. & Gō, N. (1982) *Nature (London)* **296**, 776–778.
15. Deisenhofer, J. & Steigemann, W. (1975) *Acta Crystallogr. Sect. B* **31**, 238–250.
16. Hartmann, H., Parak, F., Steigemann, W., Petsko, G. A., Ponzi, D. R. & Frauenfelder, H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4967–4971.
17. Levy, R. M., Karplus, M. & McCammon, J. A. (1981) *J. Am. Chem. Soc.* **103**, 994–996.
18. Hoch, J. C., Dobson, C. M. & Karplus, M. (1982) *Biochemistry* **21**, 1115–1125.
19. Levy, R. M. & Szabo, A. (1982) *J. Am. Chem. Soc.* **104**, 2073–2074.
20. Ichiye, T. & Karplus, M. (1983) *Biochemistry* **22**, 2884–2893.
21. Levitt, M. (1981) *Nature (London)* **294**, 379–380.
22. Karplus, M. & Kushick, J. N. (1981) *Macromolecules* **14**, 325–332.
23. Fanconi, B. & Finegold, L. (1975) *Science* **190**, 458–461.
24. Hutchens, J. O., Cole, A. G. & Stout, J. W. (1969) *J. Biol. Chem.* **244**, 26–35.
25. Privalov, P. L. (1979) *Adv. Protein Chem.* **33**, 1–71.
26. Sturtevant, J. M. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 2236–2240.
27. Jacrot, B., Cusack, S., Dianoux, A. J. & Engelman, D. M. (1982) *Nature (London)* **300**, 84–86.