Biochemistry

 Volume 21, Number 10 May 11, 1982

Protein Dynamics in Solution and in a Crystalline Environment: A Molecular Dynamics Study†

W. F. van Gunsteren‡ and M. Karplus*

ABSTRACT: The effect of a solvent and a crystalline environment on the dynamics of proteins is investigated by the method of computer simulation. Three 25-ps molecular dynamics simulations at 300 K of the bovine pancreatic trypsin inhibitor (BPTI), consisting of 454 heavy atoms, are compared: one of BPTI in vacuo, one of BPTI in a box with 2647 spherical nonpolar solvent atoms, and one of BPTI surrounded by fixed crystal image atoms. Both average and time-dependent molecular properties are examined to determine the effect of the environment on the behavior of the protein. The dynamics of BPTI in solution or in the crystal environment are found to be very similar to that found in the vacuum calculation. The primary difference in the average properties is that the equilibrium structure in the presence of solvent or the crystal field is significantly closer to the X-ray structure than is the vacuum result; concomitantly, the more realistic environment leads to a number density closer to experiment. The presence of solvent has a negligible effect on the overall magnitude of the positional or dihedral angle fluctuations in the interior of the protein; however, there are changes in the decay times of the fluctuations of interior atoms. For surface residues, both the magnitude and the time course of the motions are significantly altered by the solvent. There tends to be an increase in the displacements of long side chains and the flexible parts of the main chain that protrude into the solvent. Further, these motions tend to have a more diffusive character with longer relaxation times than in vacuo. The crystal environment has a specific effect on a number of side chains which are held in relatively fixed positions through hydrogen-bond and electric interactions with the neighboring protein atoms. Most of the effects of the solution environment seem to be sufficiently nonspecific that it may be possible to model them by applying a mean field and stochastic dynamic methods.

Over the last few years the internal dynamics of globular proteins have become a subject of intense interest. Both the experimental and the theoretical work in the field have been reviewed recently (Karplus & McCammon, 1981). Experimental methods to probe the protein dynamics include X-ray diffraction, fluorescence depolarization, fluorescence and phosphorescence quenching, hydrogen-exchange measurements, nuclear magnetic resonance, and photoexcitation experiments (Karplus & McCammon, 1981; Matthews, 1976; Yu, 1977; Shepherd, 1975; Rigler, 1977; Lakoziew & Weber, 1973; Saviotti & Galley, 1974; Woodward & Hilton, 1979; Campbell, 1977). As to the theoretical aspects, three different types of methods are being used to investigate the dynamical properties of proteins. First, approximate dynamical information can be obtained by calculating effective potential surfaces for specific processes, such as allosteric transitions (Gelin & Karplus, 1977) and aromatic ring and other side-chain rotations (Gelin & Karplus, 1975, 1979; Hetzel et al., 1976). This method is of a static character and so provides only partial information on the dynamics. A second approach is to perform a normal-mode vibrational analysis for a protein. The weakness of the method is that it ignores the anharmonic properties of the molecule and that it requires diagonalization of a matrix of dimensionality equal to the number of degrees of freedom. This method has so far been applied only to segments of proteins, such as α helices and β sheets (Shimanouchi, 1976; Rabolt et al., 1977; Levy & Karplus, 1979). The third and most powerful method is molecular dynamics (MD), in which the classical equations of motion for all of the atoms of the molecule are integrated for a finite period of time. The resulting trajectory for the molecule, which is assumed to provide an adequate coverage of the phase space of the system in equilibrium at a given temperature, is used to compute the average and time-dependent properties of the system. The method has been applied to the dynamics of isolated proteins, such as the bovine pancreatic trypsin inhibitor (BPTI) (McCammon et al., 1977, 1979; Karplus & McCammon, 1979), and to the study of activated processes in proteins (McCammon & Karplus, 1979).

Although these theoretical studies have made an essential contribution to our understanding of dynamical processes in proteins on the atomic level, it is important to determine

† From the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138. *Received March 9, 1981.* Supported in part by the National Science Foundation and the National Institutes of Health. W.F.v.G. had a fellowship from the Z.W.O. (Netherlands).

‡ Present address: Department of Physical Chemistry, State University of Groningen, Nyenborgh 16, 9747 AG, Groningen, The Netherlands.

0006-2960/82/0421-2259$01.25/0

whether the approximations employed in the calculations have introduced any significant errors into the results. This seems unlikely with respect to the qualitative features of the dynamics but may be more important for quantitative studies and detailed comparisons with experiment. An obvious element that has been neglected so far is the interaction of the protein under study with the crystal or solvent environment. Because of the role of the solvent in stabilizing the native structure (Kauzmann, 1959), the validity of dynamic simulations of proteins in vacuo has been questioned. The comparison of a dipeptide simulation in vacuo and in water (Rossky & Karplus, 1979) shows that the short-time dynamics (on the order of a few picoseconds) are little affected by solvent. However, the much greater size of a protein and the resulting increase in magnitude and complexity of its motions make it essential to have direct information on the effect of solvent. This is obtained most simply by means of computer simulations including solvent since experimental studies of the dynamics of nonsolvated proteins are difficult; examination of dried powders by techniques such as solid-state NMR would be of considerable interest in this regard, though their comparison with vacuum calculations would not necessarily be appropriate.

Differences between the dynamic results obtained from vacuum and solution simulations or experiments can originate from two primary sources. The first involves effects on the equilibrium behavior of the protein and the second direct effects on the dynamics. Clearly, the presence of solvent and the interactions between it and the protein could alter the equilibrium structure, which in turn could perturb the magnitude and time course of positional and other fluctuations. Such effects can be described as originating in the solvent contribution to the potential of mean force (McQuarrie, 1973) between protein atoms. In addition, there are the purely dynamic effects due to the solvent. The presence of solvent results in a frictional drag (finite viscosity) and collisional impulses (Brownian random force) that act on the atoms of the protein. These can alter the time dependence of the motions but not their magnitude. Given these various possible effects, it is the primary purpose of this paper to use a molecular dynamics simulation to investigate the consequences of solvation on the equilibrium and dynamic properties of a protein. An effect of solvent on protein dynamics has been investigated experimentally by Frauenfelder and co-workers (Beece et al., 1980) in their detailed study of the viscosity dependence of carbon monoxide rebinding to myoglobin.

Solvation studies have been performed for small solutes, like spheres (Owicki & Scheraga, 1977; Geiger et al., 1979; Pangali et al., 1979), NH_4Cl (Bishop et al., 1979), a dipeptide (Rossky & Karplus, 1979), or short polymers (Weber & Helfand, 1979). For larger solutes like proteins, no solution simulations exist, except for those in which the protein solute was kept rigid and only the solvent molecules were studied (Hermans & Rahman, 1976; Hagler & Moult, 1978). In the present study both the protein and the solvent are included in the molecular dynamics calculation. Since our primary concern is with the effects of the solvent on the dynamics of the protein, we use the simplest possible model for the solvent molecules; that is, we represent them by neutral spherical atoms that interact with each other and with the protein by a van der Waals force. This means that the hydrogen-bond interactions that would be present if the protein were in aqueous solution are being neglected; the use of a realistic water model, like the ST2 model (Stillinger & Rahman, 1972), is desirable but would have made the simulation too time consuming . Further, it is likely that certain important properties of the solvent—namely, its attractive interactions with the protein atoms and the excluded volume effects due to its presence—are satisfactorily represented by the van der Waals particles used in the solvent simulation. This conclusion is in accord with current theories for the statistical mechanics of simple liquids (Chandler, 1974). We expect, therefore, that the molecular dynamics simulation of the protein, the bovine pancreatic trypsin inhibitor (BPTI, 454 heavy atoms), in a bath of about 2500 spherical solvent atoms provides a satisfactory first approximation of the solvent effect on the dynamics of proteins. Further, when simulations with more realistic water models are made on faster computers, the present study will provide a basis for comparison; that is, it will make it possible to separate the effects of directional hydrogen bonds and water structure from the more general solvation properties examined in this simulation.

In addition, to the study of proteins in solution, it is important to determine the effect of the crystalline environment on protein structure and dynamics. This is of particular interest at present because of recent attempts to extract dynamical information from the temperature factors determined in X-ray analyses of crystal structures (Frauenfelder et al., 1979; Artymiuk et al., 1979; Northrup et al., 1980a,b). Some experimental studies have been made to determine whether the specific crystalline environment of a protein has any effect on its structural and dynamic properties (Wagner & Wüthrich, 1978). From a theoretical point of view, the simplest way to model the crystal environment is to surround the protein molecule that is the subject of the dynamics simulation by fixed crystal image atoms; that is, a constant force field originating from the crystal environment is applied to the molecule of interest. Such a model does not include the effect of the solvent molecules that are present in the crystal nor of the motion of the atoms in neighboring protein molecules. Nevertheless, we expect this type of simulation, which we have performed for BPTI, to provide preliminary information on the perturbations of the structure and dynamics by the crystalline environment. It has been shown in a previous study (Gelin & Karplus, 1979) of the static potential for surface side-chain rotations in BPTI that the present model for the crystal field yields satisfactory structural results.

This paper compares the results of three molecular dynamic runs on BPTI. The first is a simulation of an isolated BPTI molecule (LC), the second treats one BPTI molecule in a rectangular box of 2647 solvent atoms (LCS), and the third one is of a BPTI molecule surrounded by fixed crystal image atoms from its protein neighbors (LCC). The model for the protein, the solvent, and the crystal environment and the computational procedure are described. The simulation of BPTI in solution is compared to that in vacuo, and the simulation of BPTI in the crystal environment is compared to the other two molecular dynamic runs. The discussion and conclusions follow.

Model and Computational Procedure

The properties of the model that is used to simulate BPTI (58 amino acid residues, 454 heavy atoms) plus four internal hydrogen-bonded water molecules were described previously (Gelin & Karplus, 1979; McCammon et al., 1979). Consequently, we present here only the aspects of the model by which the solvent and crystal environments are introduced.

In all three runs (LC, LCS, and LCC), the equations of motion were integrated in Cartesian coordinates with the Verlet algorithm (Verlet, 1967). The bond lengths were constrained by applying the Shake method with a relative tolerance of 10^{-4} (Ryckaert et al., 1977; van Gunsteren & Berendsen, 1977). No metric tensor potential has been in-

cluded in the interaction, since for bond-length constraints alone, the effect is unimportant (van Gunsteren, 1980). When bond-length constraints are employed, the MD time step Δt can be taken as large as $\Delta t = 0.004$ "ps" without altering the dynamics of the protein (W. F. van Gunsteren and M. Karplus, unpublished results). Throughout this paper we use the following units: energy in kilocalories per mole, mass in atomic mass units, μ, length in angstroms, time in "ps" = $2/(4.184)^{1/2}$ ps = 0.98 ps, and temperature in kelvin, unless it is explicitly stated otherwise.

BPTI in a Box with Solvent Atoms. The solvent is modeled by extended atoms of atom type OH2 (McCammon et al., 1979), having mass = 18.0154 μ and van der Waals radius R_{vdW} = 1.70 Å. The solvent atoms have no charge or dipole moments and so interact only through the van der Waals term in the interaction function potential (Gelin & Karplus, 1979; McCammon et al., 1979). For the OH2–OH2 interaction, the well depth is ϵ = 0.1416 kcal mol^{-1} and the radius σ = 3.029 Å. This is to be compared with ST2 water for which the Lennard–Jones O–O interaction is ϵ = 0.07575 kcal mol^{-1} and σ = 3.10 Å (Stillinger & Rahman, 1972).

The starting configuration for the LCS run was obtained by first placing solvent atoms on a cubic lattice in a rectangular box. The smallest distance between the solvent atoms was chosen to be 3.1034 Å in order to obtain a mass density of 1 g cm^{-3}. The dimensions of the box (40.3442, 40.3442, and 52.7578 Å) were selected such that all protein atoms would be separated by more than 6 Å from the walls when the protein is placed in the center of the box and such that they would be a multiple of the OH2–OH2 distance of 3.1034 Å. The rectangular arrangement was randomized by adding to each Cartesian coordinate of each atom a random number uniformly distributed between ±0.31034 Å. Then, the protein with the X-ray geometry (Deisenhofer & Steigemann, 1977, as obtained from Brookhaven Protein Data Bank, Brookhaven National Laboratory, Upton) was placed in the center of the box, and all solvent atoms in lattice cells occupied by the protein atoms were removed. The resulting system consisted of 3105 atoms (458 BPTI, 2647 solvent) and had a mass density of 1.05 g cm^{-3}; the density of the solvent is 0.99 g cm^{-3} (Table I). The four water molecules that are hydrogen bonded within BPTI are treated as part of the protein. All protein atoms are farther from the walls of the box than 6.21 Å. This means that, with the periodic boundary conditions, atoms of proteins in adjacent boxes are separated by at least four layers of solvent atoms. For minimization of the artifacts arising from the application of periodic boundary conditions, the long-range electrostatic interactions between atoms in neighbor proteins were omitted; i.e., only the protein–protein image van der Waals interactions were included in the potential function. However, since atoms from neighbor proteins were more than 10 Å apart at all times during the run, the protein–protein interaction term was equal to zero. As in the LC run, a spherical cutoff was applied in the interaction potential, but in the LCS run different cutoff radii, R_c, and switching function radii, R_s, were used for the electrostatic (R_s = 7.5 Å, R_c = 8.0 Å) and the van der Waals (R_s = 7.0 Å, R_c = 7.5 Å) interaction for reasons of computational efficiency. Unlike the LC run, no nonbonded pair lists were used. Instead the interacting pairs of atoms were determined at every (MD or energy minimization) step by means of a pair-search scheme similar to that of Quentrec & Brot (1973). For such a large system, this scheme turned out to be a factor of 1.5–2 faster than the linked list (Hockney et al., 1974) or pair list (Verlet, 1967) techniques. The required computer time for the LCS run amounts to about 19 h CPU time per ps on a VAX 11-780.

For equilibration of the protein, the large repulsive energy of the starting configuration of BPTI plus solvent (Table I) was reduced by performing an energy minimization (EM), using the steepest-descent algorithm for constrained energy minimization (van Gunsteren & Karplus, 1980). Although the total potential energy was still dropping by 1–5 kcal mol^{-1} per step after 100 EM steps, no further minimization was performed since the various energies had decreased to reasonable values (Table I); the steepest-descents step size was as small as 0.0065 Å, and the root mean square gradient (mean over all atoms) of the potential along the nonconstrained degrees of freedom amounted to 3.1 kcal mol^{-1} Å^{-1}. Atomic velocities were taken from a Maxwellian distribution with a temperature of 450 K to start the dynamics run. The reference temperature for the LCS run was taken equal to that of the LC run (T_0 = 300 K). During the first "picoseconds" of the run, the velocities were rescaled only 3 times to keep the temperature within ΔT_0 = 10 K of T_0. No further velocity rescaling was required for the remainder of the 33-"ps" LCS run. After 5 "ps", the total potential energy appeared to be approaching an asymptotic value (Table I), so the equilibration period was limited to the relatively short period of 8 "ps"; this was necessary because of the time-consuming nature of the 3105 particle MD run. The final 25 "ps" of the run was used for analysis. We note that even after 33 "ps" the electrostatic energy of the protein is still dropping more or less monotonically; this is true also for he LC run.

Although the solvent atoms have a mass equal to that of a water molecule, their diffusion constant D is about 4 times as high as that of water (D = 0.24 Å^2/"ps" at 303 K), since the solvent atoms lack a molecular structure and cannot form hydrogen bonds. This should be kept in mind when the time scale of processes in the protein that depend on the presence of the solvent is considered. The mobility of the solvent atoms increases as a function of their distance from the protein surface (that is, from the nearest protein atom), especially across the first two layers of solvent atoms around the protein; i.e., the average value of D is 0.92 Å^2/"ps", while the first and second layers have values of 0.68 and 0.83 Å^2/"ps", respectively.

BPTI Surrounded by Fixed Crystal Protein Image Atoms. The orthorhombic BPTI crystal used for the X-ray structure determination (Deisenhofer & Steigemann, 1977, as obtained from Brookhaven Protein Data Bank, Brookhaven National Laboratory, Upton) has lattice parameters a = 43.1 Å, b = 22.9 Å, and c = 48.6 Å, space group $P2_12_12_1$ and one molecule per asymmetric unit (Huber et al., 1970). The coordinates of the crystal image atoms which form the crystalline environment of the central molecule can be obtained from the X-ray coordinates of the central molecule by performing the symmetry transformation corresponding to the $P2_12_12_1$ space group (Henry & Lonsdale, 1965). In this way 2190 crystal image atoms were found to be present in a box determined by the planes x = −7.24 Å, x = 35.76 Å, y = 2.99 Å, y = 40.52 Å, z = −15.00 Å, and z = 27.66 Å, which surrounds the central molecule. The size and position of this box have been chosen such that all atoms of the central molecule are farther than 5 Å from one of the walls. We note that the coordinate system used here (LCC run) is identical with that of Deisenhofer & Steigemann (1977, as obtained from Brookhaven Protein Data Bank, Brookhaven National Laboratory, Upton), but differs from the one (LCS run) used in the previous section. The positions of the 2190 crystal image atoms were kept fixed during the EM and MD runs; their role in the simulation is

Table I: Average Temperature, Energies, and Solvent Density as a Function of Time[a]

LCS	time period ("ps")	$\langle T \rangle$	$\langle E_{kin} \rangle$	$\langle T^p \rangle$	$\langle T^s \rangle$	$\langle E_{pot} \rangle$	$\langle E_\theta{}^p \rangle$	$\langle E_\xi{}^p \rangle$	$\langle E_\phi{}^p \rangle$	$\langle E_{hb}{}^p \rangle$	$\langle E_{el}{}^p \rangle$	$\langle E_{vdw}{}^p \rangle$	$\langle E_{vdw}{}^{p-s} \rangle$	$\langle E_{vdw}{}^s \rangle$	ρ^s
X-ray						30000	106.3	6.7	67.7	-119.8	-34.6	-177.8	30000	-426.5	0.994
EM						-3025	93.0	46.4	70.8	-128.2	-75.7	-273.4	-354.7	-2403	0.994
MD	4-5	303	2660	306	302	-2568	198.5	52.7	84.2	-119.1	-225.9	-290.5	-373.1	-1895	0.993
	7-8	302	2658	303	302	-2570	197.5	51.1	85.6	-121.9	-248.3	-292.1	-366.0	-1876	0.990
	12-13	303	2660	309	302	-2579	201.2	47.6	90.0	-118.0	-252.5	-295.0	-367.6	-1884	0.997
	17-18	304	2676	323	302	-2599	203.1	46.6	87.8	-121.3	-258.5	-299.3	-370.3	-1887	0.991
	22-23	304	2674	308	304	-2603	203.7	46.6	91.9	-123.0	-278.8	-304.2	-366.7	-1872	1.000
	27-28	305	2681	304	305	-2614	201.8	45.4	93.1	-121.1	-287.2	-310.5	-366.9	-1869	0.986
	32-33	304	2673	301	304	-2610	207.5	44.7	94.9	-120.1	-294.1	-314.1	-358.6	-1871	0.985
LCC													$\langle E_{hb}{}^{p-c} \rangle$	$\langle E_{el}{}^{p-c} \rangle$	$\langle E_{vdw}{}^{p-c} \rangle$
X-ray						-229.6	106.3	6.7	67.7	-119.9	-34.6	-188.4	-9.7	-25.4	-32.4
EM						-600.4	91.7	14.1	76.6	-131.2	-157.9	-342.9	-10.4	-99.6	-41.0
MD	4-5	304	271.5	304		-468.8	201.6	41.8	95.6	-107.7	-205.8	-334.8	-8.4	-115.1	-36.0
	9-10	303	273.1	303		-476.4	199.6	44.3	93.2	-105.6	-211.8	-326.4	-7.4	-125.7	-36.6
	14-15	304	274.0	304		-483.0	203.6	43.6	95.3	-107.0	-215.4	-335.9	-6.9	-122.5	-37.6
	19-20	301	271.2	301		-475.3	199.6	40.9	95.8	-104.0	-208.1	-324.5	-6.2	-132.0	-36.6
	24-25	303	273.1	303		-482.5	203.0	42.9	95.9	-109.5	-215.9	-326.9	-6.6	-129.5	-36.0
	29-30	306	275.8	306		-488.6	205.3	42.8	96.0	-108.0	-217.5	-335.8	-6.7	-129.3	-35.4
	34-35	308	277.7	308		-493.6	198.9	41.7	94.8	-104.6	-218.4	-335.4	-6.3	-128.8	-35.5
	39-40	309	278.5	309		-495.4	205.1	44.6	95.1	-106.0	-221.0	-340.5	-6.6	-130.7	-35.4

[a] The values after energy minimization are denoted by EM. The time intervals of the MD runs over which averages are given are in "picoseconds" ("ps"). The temperature (T) is in K, and the energies are in kilocalories per mole. The energy contributions listed are the kinetic energy (E_{kin}), potential energy (E_{pot}), bond angle energy (E_θ), improper dihedral angle energy (E_ξ), dihedral angle energy (E_ϕ), hydrogen-bond energy (E_{hb}), electrostatic energy (E_{el}), and the van der Waals energy (E_{vdw}). The internal protein energies are denoted by the superscript p, solvent energies by the superscript s, protein–solvent energies by p–s, and protein–crystal energies by p–c. Since the crystal image atoms are kept fixed, only half of the protein–crystal energy has been listed. The symbol ρ^s denotes the solvent number density. It is calculated by counting the number of solvent atoms in the rectangular layers directly adjacent to the walls of the box. It is not a time average but calculated from the last solvent configuration of each period.

to supply a constant external force field. There are 23 central molecule, crystal image atom pairs separated by less than 3.5 Å in the X-ray configuration and 120 pairs out to 4.0 Å; Lys-15, Arg-17, and Arg-29 have the most protein neighbor contacts.

In the LCC run, the nonbonded interactions within the central molecule are treated in exactly the same way as for the LC run; that is, the same switching function radius R_s = 7.5 Å and cutoff radius R_c = 8.0 Å were applied for the van der Waals as well as the electrostatic interactions. When making the nonbonded pair list, a cutoff radius R_l = 8.5 Å was used, which allows us to update it only every 0.02 "ps" (5 steps). For the nonbonded interactions between the central molecule and the crystal image atoms, the shielding effect of the water molecules between neighbor molecules in the crystal was taken into account in a simplified way by choosing a shorter ranged interaction, that is, R_s = 4.0 Å and R_c = 5.0 Å for central molecule, crystal image nonbonded pairs. The number of central molecule, crystal image atom pairs is very large (about 10^6), but only about 10^3 are at a distance smaller than the cutoff radius (R_c = 5.0 Å) and contribute to the interaction. The most important terms arise from electrostatic interactions between charged side chains. It is computationally efficient, therefore, to use a relatively large cutoff radius (R_l = 8.0 Å) when the pair list is made. Combined with the immobility of the crystal image atoms, this enables us to update the pair list only every 0.12 "ps" (30 steps). Within the central molecule, the total number of nonbonded pairs equals about 10^5, of which about 2×10^4 are less than R_c = 8.0 Å apart. The required computer time for the LCC run is about 46 min CPU time per ps (LC: 37 min).

The central molecule, crystal image hydrogen-bonded pairs were obtained by checking whether the donor–acceptor pairs satisfy the criteria for hydrogen bonding (W. F. van Gunsteren and M. Karplus, unpublished results). Only four donor–acceptor pairs obeyed all criteria; they are Asp-3 N–Glu-49 O_{ϵ_1}, Arg-39 N_ϵ–Glu-49 O_{ϵ_1}, Arg-39 N_{η_1}–Glu-49 O_{ϵ_2}, and Arg-39 N_{η_2}–Tyr-21 O_η.

A corresponding equilibration procedure as for the LCS run was employed. The strain was removed from the starting configuration by performing constrained conjugate gradient energy minimization with a step size of 1.5 Å (van Gunsteren & Karplus, 1980). After 100 steps, the total potential energy dropped less than 1 kcal mol^{-1} per EM step and the root mean square (rms) gradient of the potential along the nonconstrained degrees of freedom amounted to 1.4 kcal mol^{-1} Å^{-1}. With the resulting atom positions and the velocities chosen from a Maxwellian distribution for 450 K, the MD run was started. The reference temperature was taken equal to that of the LC and LCS runs (T_0 = 300 K). During the first 8 "ps" of the LCC run, the velocities were rescaled 6 times in order to keep the temperature within ΔT_0 = 8 K of T_0. No more rescaling was required for the remainder of the 40-"ps" LCC run. After 10 "ps", the potential energy leveled off (Table I); the equilibration period was terminated at 15 "ps". The final 25 "ps" of the run was used for analysis. We note that the difference between the van der Waals energy of the X-ray structure for the LCS run and the LCC run (Table I) originates from the different switching function and cutoff radii, R_s and R_c, that were used in these runs.

Computation of Properties. All averages and correlation functions reported in the next sections are calculated from the final 25-"ps" portions of the runs; they make use of coordinates, velocities, or energies that were stored every 0.008 "ps". In addition to averages over the total analysis period of 25 "ps", averages over 5 "ps" have also been determined to bring out the dependence of the results on the time period involved. In the LCS run, the protein shows an overall translational and rotational (Brownian) motion. The LC run has no overall motion since the total linear momentum and angular momentum have been set equal to zero by adjustment of the initial velocities, and no overall motion is generated during the run because of the vacuum boundary condition. The LCC run shows some overall motion of the central molecule, but it is considerably restricted by the fixed crystalline environment of the protein. For elimination of the overall motion of the protein when a time series of protein configurations is averaged, the first configuration of the series is taken as the reference configuration. For all the subsequent configurations, the centers of mass are made identical with that of the first one, and each is rotated around its center of mass such that the mean square difference of the C_α atom coordinates with respect to the first configuration is a minimum (McLachlan, 1979). When a calculation is made with the motion of and about the center of mass eliminated in this way, it is denoted by the symbol cm; otherwise no symbol is used.

Molecular Dynamics of BPTI in Solution (LCS)

In this section the effects of solvent on the dynamical properties of a protein in solution are investigated by comparing the results of the LCS run for BPTI in a box with solvent atoms with those of the LC run of BPTI under vacuum. The average temperatures of the two runs are nearly the same (305.2 K for LC and 305.7 K for LCS). However, the solvent run shows larger fluctuations of the protein temperature because its kinetic energy is coupled to that of the solvent [$\langle(\Delta T)^2\rangle^{1/2}$ is 10.6 K for LC and 13.6 K for LCS]. The classical protein heat capacity could not be calculated from the LCS run, since the series of protein configurations does not form a microcanonical ensemble (as is the case for the vacuum run), nor does it form a canonical ensemble since the protein temperature is not constant. For the protein plus the solvent, the temperature and its fluctuation averaged over 25 "ps" are $\langle T\rangle$ = 304.5 K and $\langle(\Delta T)^2\rangle^{1/2}$ = 2.8 K; this yields an overall heat capacity of $c_{at} = 2.32\ k_B$ where k_B is the Boltzmann constant. The presence of bond-length constraints in the protein can be accounted for in an approximate way by adding $3k_B$ per constraint to the total heat capacity, since unconstrained bond lengths are essentially harmonic. With this correction, we find $c_{at} = 2.77k_B$ for the system of protein plus solvent. This is to be compared with the bond length constraint corrected LC protein heat capacity of $3.27k_B$. It is clear from this result that the heat capacity of the solvent is significantly smaller than that of the protein. This is as expected for liquids made up of van der Waals atoms. Further, although the van der Waals nonbonded interactions between protein atoms play an important role, the classical heat capacity of the protein is increased over that expected for a system of van der Waals atoms by the harmonic character of the bonds and bond angles. Quantum corrections to such classical heat capacity results have been discussed previously (Karplus & McCammon, 1981).

The kinetic energy is, in general, equally distributed over the backbone, as well as between backbone and side-chain atoms, as it was in the LC run. The velocity autocorrelation functions for the C_α atoms are very similar to each other and to those of the LC run. For atoms in exposed side chains, they are less oscillatory and more damped relative to interior atoms (see Figure 1). However, the presence of the solvent makes little difference; the correlation function falls within 0.1 "ps" to negative values and after 0.2 "ps" oscillates around zero.

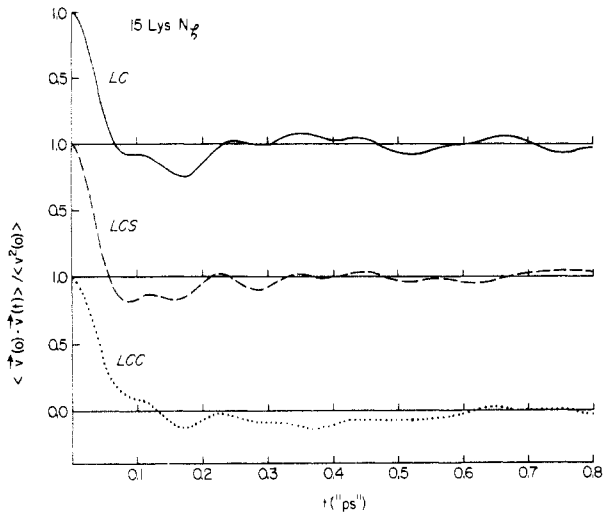

FIGURE 1: Velocity autocorrelation functions of Lys-51 N_ζ in the three runs.

Table II: Root Mean Square Fluctuations $\overline{\langle(\Delta x)^2\rangle}^{1/2}$ of the Atom Cartesian Coordinates (Å)

	Comparison of Time Averages[a]					
	all atoms (458)			C_α atoms (58)		
period ("ps")	LC	LCS (cm)	LCC (cm)	LC	LCS (cm)	LCC (cm)
0–5	0.55	0.55	0.53	0.43	0.38	0.40
5–10	0.56	0.51	0.50	0.41	0.37	0.36
10–15	0.57	0.47	0.48	0.40	0.34	0.35
15–20	0.56	0.45	0.57	0.41	0.35	0.45
20–25	0.56	0.51	0.52	0.39	0.39	0.39
rms*	0.56	0.50	0.52	0.41	0.37	0.39
0–25	0.72	0.78	0.67	0.54	0.54	0.52
X-ray	0.63			0.58		
ref 1[c]	0.90			0.74		
ref 2	0.75			0.60		

	Comparison of Time Averages[a]					
	C_{ϵ_1} atoms (8)			N_ζ atoms (4)		
period ("ps")	LC	LCS (cm)	LCC (cm)	LC	LCS (cm)	LCC (cm)
0–5	0.56	0.64	0.65	0.61	1.02	0.69
5–10	0.67	0.59	0.63	0.64	0.70	0.87
10–15	0.69	0.56	0.60	0.81	0.68	0.52
15–20	0.60	0.50	0.77	0.73	0.66	0.78
20–25	0.76	0.54	0.71	0.63	0.73	0.73
rms*	0.66	0.57	0.67	0.68	0.77	0.73
0–25	0.82	0.73	0.94	0.88	1.40	1.00
X-ray	0.56			0.96		

	Comparison of Atoms[b]						
atom name	LC	LCS (cm)	LCC (cm)	atom name	LC	LCS (cm)	LCC (cm)
all	0.72	0.78	0.67	C_α	0.54	0.54	0.52
N	0.52	0.50	0.50	C_β	0.62	0.69	0.62
C	0.53	0.51	0.50	C_γ	0.74	0.89	0.71
O	0.66	0.65	0.67	C_δ	0.78	1.15	0.77

[a] The symbol rms* denotes the root mean square of the five "picosecond" values. The number of atoms with a certain atom name included in the average is given in parentheses. [b] Averages over 25 ps. [c] References: (1) McCammon et al. (1977); (2) Karplus & McCammon (1979).

The oscillation observed in the velocity autocorrelation functions, particularly for interior atoms, is in accord with expectations for a dense fluid.

The root mean square fluctuations, $\overline{\langle(\Delta x)^2\rangle}^{1/2}$, of the atom positions are given in Table II; both 25- and 5-"ps" averages are given. The bar denotes a sum over Cartesian components

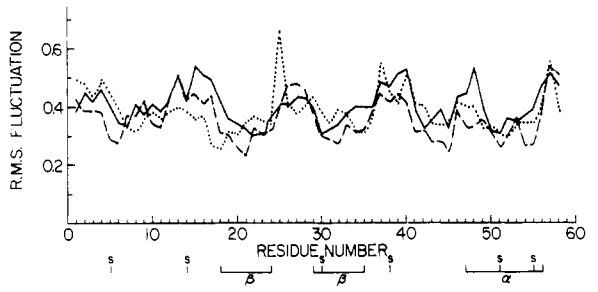

FIGURE 2: Root mean square fluctuations of C_α atoms for the three runs; time average over 5 "ps"; (—) LC, (---) LCS, (···) LCC.

(if possible) and subsequently averaging over atoms with a given name. The largest difference between the solution (LCS) and vacuum (LC) values is found for atoms at the end of side chains that protrude into the solvent, like the N_ζ atoms in the Lys residues. They appear to move about more in the LCS run. The values for the C_{ϵ_1} atoms indicate that the aromatic rings have slightly less space to move in solvent than in vacuo. Averaged over all atoms, the solvent run shows slightly larger fluctuations over a period of 25 "ps", but smaller ones over a period of 5 "ps". The observed differences may result from a greater "drift" in the structure of the protein, particularly of the exterior side chains, during the LCS run relative to the LC run. That some long-time motional component is present in both runs is clear from the fact that the 5-"ps" averages yield considerably smaller values for the fluctuations than the 25-"ps" averages (see Table II). The comparison of the fluctuations by atom name in Table II makes clear that the backbone atoms (N, C, C_α, and O) have the same behavior in the LCS and LC runs, with the carbonyl oxygen (O) having slightly larger fluctuations than N, C, and C_α. As one goes out along the side chains, there is a progressive increase in the magnitude of the fluctuations, with the values from the solvent (LCS) run increasing more rapidly than those in the vacuum (LC) run.

The rms fluctuations from the 5-"ps" averages for the C_α atoms are displayed in Figure 2. Although the LCS and LC results are similar, quantitative differences between the runs are seen in the atom-by-atom comparison. In the results for LC and LCS, the ends of the protein chain and the extended loop (residues 25–28), which form the most flexible parts of the molecule, have large fluctuations. By contrast, the regions with secondary structure (β sheet and α helix) have low fluctuations. There seems to be no correlation between the location of the disulfide bonds and the magnitude of the fluctuations. In Figure 3 the rms fluctuations, $\overline{\langle(\Delta x)^2\rangle}^{1/2} = [3B/(8\pi^2)]^{1/2}$, are plotted, as they are obtained from the X-ray temperature (B) factors; Figure 3a shows the C_α results and Figure 3b those for the side chains; the latter are obtained by averaging the mean square fluctuations of the individual side-chain atoms obtained from their B values. The experimental results are on the whole somewhat larger than those calculated from the dynamics. This suggests that at least part of the observed temperature factors originate in crystal disorder rather than dynamic fluctuations (Frauenfelder et al., 1979; Artymiuk et al., 1979; Northrup et al., 1980a,b). There seems to be a weak correlation between the calculated and the X-ray C_α fluctuations as a function of residue number; particularly for the first 20 or so residues, the calculations and the experimental data agree, including the low fluctuations in the β-sheet region. The 5-"ps" rms fluctuations of the side chains are displayed in Figure 4 together with the percentage of each residue that is exposed to the solvent. Although the LC and LCS curves exhibit the same features, the latter curve is

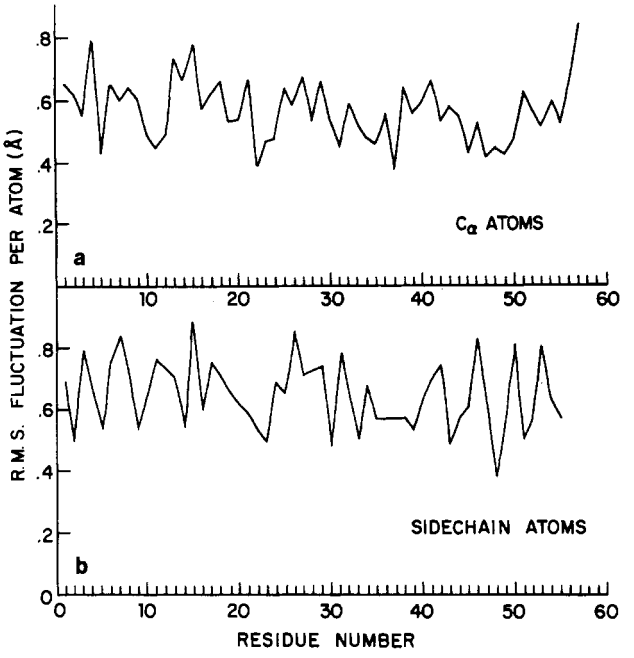

FIGURE 3: Root mean square fluctuations derived from X-ray temperature factors (J. O. Deisenhofer, private communication): (a) C_α atoms; (b) side-chain atoms.

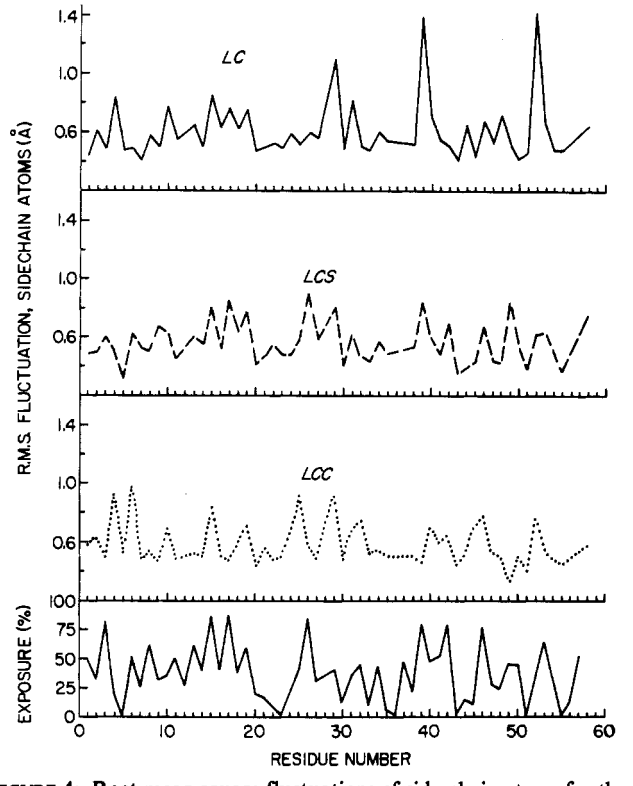

FIGURE 4: Root mean square fluctuations of side-chain atoms for the three runs: time average over 5 "ps". The side-chain exposure is also given.

significantly closer to that for the exposure of the residues than is the former. There is a clear positive correlation between the LCS curve (Figure 4) and the X-ray curve in Figure 3b. That the charged polar side chains show significantly larger displacement, in general, than the other side chains is in agreement with experiment (Frauenfelder et al., 1979; Northrup et al., 1980a,b); this is due primarily to the fact that they stick out into solution in most cases. The side chains of Arg-39 and Glu-49 show much larger fluctuations in the LCS calculation than in the X-ray results; the very small experimental value is due to the hydrogen bond between the two residues in neighboring molecules of the crystal (see above). The displacements of the charged side chains are about the same in vacuo and in the solvent, when measured over 5 "ps". For the neutral and nonpolar residues, the LCS displacements are generally smaller than those found in the LC run. When the 5- and 25-"ps" values are close to each other in a given run, the sampling of coordinates in the run may be assumed to give a complete picture of the short time fluctuations. For these cases (Ser, Thr, and Val), the LCS and LC fluctuations turn out to be very similar. Certain differences between the LCS and LC values can be explained by specific details of the two runs; e.g., the Cys values averaged over 5 "ps" are close to each other, but over 25 "ps" the LC value is larger than the LCS values. This is caused by a transition in the Cys-30–Cys-51 disulfide bridge in the LC run.

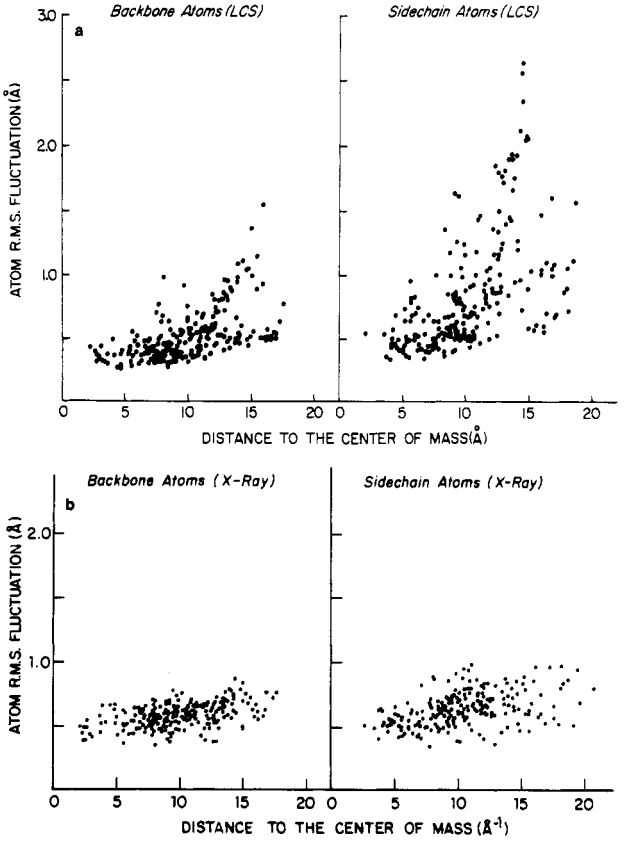

FIGURE 5: Atom rms fluctuations calculated over 25 "ps" as a function of distance from center of mass: (a) solvent run (LCS); (b) derived from X-ray temperature factors.

In Figure 5a the 25-"ps" atom rms fluctuations for the solvent run are plotted vs. the distance (r_{CM}) of the atom from the center of mass, calculated from the time-averaged structure. The results obtained for the vacuum run are very similar. Both for the backbone and the side chains there is a general tendency of the average fluctuations to increase with distance from the center of mass. This is due primarily to the fact that at larger r_{CM} values more of the atoms are on the surface of the molecule, although there is a weak dependence on r_{CM} even in the protein interior; the minimum fluctuation values are essentially independent of r_{CM} for both backbone and side-chain atoms. The present results are in accord with the analysis of temperature factors for lysozyme (Artymiuk et al., 1979) and ferrocytochrome *c* (Northrup et al., 1980a,b). The X-ray values (Figure 5b) show a smaller variation and much less dependence on r_{CM}, particularly for the side-chain atoms. The experimental results are unlikely to be correct in view of the theoretical results and analyses for other proteins (Ar-

Table III: Anisotropic and Anharmonic Behavior by Atom Name

	Anisotropy[a]					
	5 "ps"			25 "ps"		
atom name	LC	LCS (cm)	LCC (cm)	LC	LCS (cm)	LCC (cm)
all	2.61	2.69	2.57	2.51	3.06	2.54
N	2.23	2.23	2.16	2.30	2.28	2.21
C	2.23	2.27	2.21	2.34	2.36	2.27
O	2.62	2.73	2.77	2.62	2.67	2.75
C_α	2.25	2.37	2.24	2.34	2.46	2.29
C_β	2.35	2.59	2.38	2.26	2.90	2.42
C_γ	2.63	2.85	2.67	2.64	3.44	2.57
C_{ϵ_1}	2.84	2.84	2.77	2.67	2.63	2.93
N_ζ	2.78	2.96	2.81	2.92	4.22	2.86

	Anharmonic Behavior[b]					
	$\left(\frac{\langle \Delta x^4 \rangle}{3\langle \Delta x^2 \rangle^2}\right)$		$\left(\frac{\langle \Delta y^4 \rangle}{3\langle \Delta x^2 \rangle^2}\right)$		$\left(\frac{\langle \Delta z^4 \rangle}{3\langle \Delta z^2 \rangle^2}\right)$	
atom name	LC	LCS (cm)	LC	LCS (cm)	LC	LCS (cm)
all	0.884	0.910	0.993	0.981	1.012	0.991
C_α	0.842	0.903	0.965	0.951	0.998	0.980
C_{ϵ_1}	0.874	0.847	1.272	0.929	0.990	0.913
N_ζ	0.905	0.975	1.065	0.916	1.045	1.023

[a] The anisotropy of the atom positional fluctuations is defined as the ratio of the largest (max) and the smallest (min) eigenvector (e.v.) of the time-averaged positional fluctuation tensor with components $\langle \Delta x_\alpha \Delta x_\beta \rangle$ ($\alpha, \beta = 1, 2$, and 3). The averaging over atom name is done before the ratio is computed, so that the table contains values of $[\max \text{e.v.} \overline{(\langle \Delta x_\alpha \Delta x_\beta \rangle)^2}^{1/2}] / [\min \text{e.v.} \overline{(\langle \Delta x_\alpha \Delta x_\beta \rangle)^2}^{1/2}]$. The anisotropies labeled 5 "ps" are obtained by averaging the five 5-"ps" anisotropies. [b] The values correspond to 25-"ps" averages; as indicated, the ratio is obtained for each atom, and the ratios are averaged over the set.

tymiuk et al., 1979; Northrup et al., 1980a,b). However, it is important to realize that the small size of BPTI makes it somewhat special in that it has very little "interior".

Results for the anisotropic and anharmonic character of the atom rms fluctuations are listed in Table III. For all types of atoms, independent of solvent, the anisotropy averages are such that the ratio of the largest and smallest rms displacement is greater than a factor of 2. There is a tendency toward larger anisotropies with than without solvent, particularly in the 25-"ps" averages and for atoms in the side chains. The side-chain atoms, including those in aromatic rings, have a somewhat larger anisotropy than those of the main chain. Further, in the aliphatic side chains, the anisotropy increases monotonically as one goes away from the main chain ($C_\gamma > C_\beta > C_\alpha$); for the 25-"ps" averages this effect is very marked. The occurrence of such large anisotropy values suggests that it may be important to take them into account in determining the thermal parameters by X-ray diffraction. Deviations from harmonic behavior are of intrinsic interest (Karplus & McCammon, 1979; Frauenfelder et al., 1979) and of importance for the analysis of temperature factors, where harmonic behavior is assumed. For positional fluctuations that correspond to a harmonic potential (i.e., that satisfy a Gaussian distribution), the value of $\langle \Delta x^3 \rangle$ is zero, and the value of $\langle \Delta x^4 \rangle$ is equal to $3\langle \Delta x^2 \rangle^2$. For many atoms the distributions are quite asymmetric ($\langle \Delta x^3 \rangle \neq 0$). As to the peakedness of the curve, the ratio averages listed in Table III indicate that there are significant deviations of $\langle \Delta x^4 \rangle$ from the harmonic value. For individual atoms, the range of the ratio is between 0.5 and 2, though most atoms are close to the harmonic value. There are differences in the deviations from harmonic behavior in the runs with and without solvent, but they do not show any regular trends.

Table IV: Root Mean Square Difference $\overline{[\langle x \rangle - x(\text{X-ray})]^2}^{1/2}$ between the Time-Averaged Structures and the X-ray Structure[a]

atom type[b]	LC	LCS (cm)	LCC (cm)
all	3.02	1.94	2.12
N	1.90	1.23	1.46
C	2.25	1.41	1.41
O	2.36	1.53	1.70
C_α	2.20	1.35	1.52
C_β	2.50	1.46	1.89
C_γ	2.69	1.74	2.31
C_δ	3.99	2.53	2.75

[a] See footnote *a* of Table II; the X-ray structure is denoted by x(X-ray). [b] Averages over 25 "ps".

The root mean square difference, $\overline{[\langle x \rangle - x(\text{X-ray})]^2}^{1/2}$, between the time-averaged structures and the X-ray structure is given in Table IV. Because of the short equilibration period of only 8 "ps" for the LCS run, the protein is moving away to a slight degree from the (initial) X-ray structure during the analysis period. Atoms at the ends of side chains on the surface of the protein show the largest deviations from the X-ray positions, since they can move around more or less freely and the potential differs most from that in the crystal (Gelin & Karplus, 1979); the N_ζ atoms in Lys-15, Lys-26, and Lys-41 were not located in the experimental electron density map of BPTI (Deisenhofer & Steigemann, 1977, as obtained from Brookhaven Protein Data Bank, Brookhaven National Laboratory, Upton). What is most striking is that the presence of the solvent leads to a significant improvement in the agreement between the time-averaged structure and the X-ray structure; i.e., the all atom rms deviation is reduced to 1.94 Å in the LCS run, relative to 3.02 Å in the LC run. For all atoms the results for the 5-"ps" averages are very similar to the 25-"ps" averages, indicating that structural drifts are not playing an important role in the present comparisons. The LC structure deviates strongly from the X-ray structure in the extended loop (residues 25–28) and at the carboxy terminus of the protein. This is not surprising, since these are, together with the amino-terminal end of the chain, the most flexible parts of the molecule; all atoms of Ala-58 and the C and O atoms of Gly-57 were not observed in the electron density map (Deisenhofer & Steigemann, 1977, as obtained from Brookhaven Protein Data Bank, Brookhaven National Laboratory, Upton). In the solvent simulation, these very large differences between the time-averaged and X-ray structures disappear. In addition, the LCS values are closer to the X-ray results for much of the rest of the chain; only near residues 14 and 38 involved in an S–S bond is the LC structure closer to the X-ray structure than the LCS result. If we look at the individual atom types (Table IV), we see that the LCS results are closer to the X-ray value in every case. Further, the main-chain differences (N, C, and C_α) are smallest, and there is a monotonic increase in the deviation for both the LC and LCS runs along the aliphatic side chains; the main-chain O atom deviations are seen to be slightly larger than those of the rest of the main chain. Comparing the LC and the LCS structures, we find that the latter is closer to the X-ray structure than to the vacuum simulation.

In Figure 6 the position fluctuation autocorrelation functions $[C(t)]$ are displayed for certain atoms. For both the C_α and the N_ζ atom of Lys-15, the decay of the correlation function is much slower in the presence of solvent (LCS) than in vacuo (LC); further, $C(t)$ in vacuo shows a clear oscillatory behavior that is probably not present in the solvent run, though the decay of $C(t)$ in the latter is too slow to determine what actually happens. The long time tail of the LCS functions is due

Table V: Atom Fluctuation Relaxation Times

residue	atom	τ ("ps") LC	τ ("ps") LCS	τ ("ps") LCC	atom	τ ("ps") LC	τ ("ps") LCS	τ ("ps") LCC	exposure (%)
Lys-15	C_α	0.96	2.47	0.71	N_ζ	0.51	1.09	0.58	87
Lys-26	C_α	0.57	1.54	0.84	N_ζ	0.43	1.33	0.17	84
Phe-33	C_α	0.59	1.10	0.76	C_{ϵ_1}	0.35	0.84	0.56	10
Gly-36	C_α	0.27	1.25	0.37					2
Asn-43	C_α	0.45	0.89	0.78	O_{δ_1}	0.19	0.50	0.68	0
Cys-51	C_α	0.32	0.81	0.56	S_γ	0.52	0.60	0.58	0

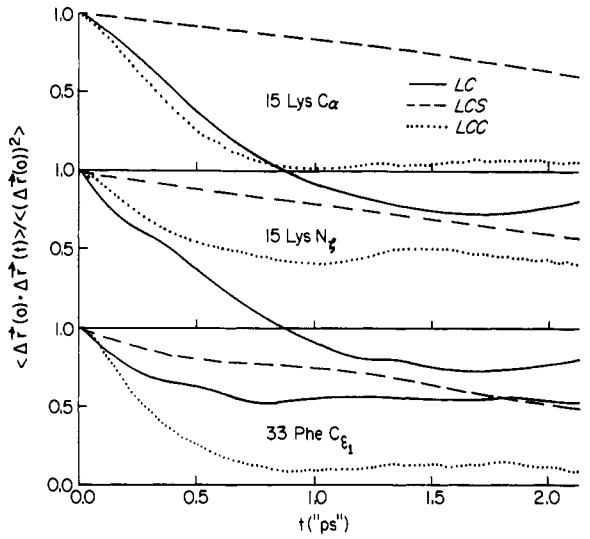

FIGURE 6: Positional fluctuation autocorrelation functions for Lys-15 C_α, Lys-15 N_ζ, and Phe-33 C_{ϵ_1} in the three runs.

to the more diffusive behavior produced by the presence of the solvent. The surface atoms are most likely to be strongly affected, though some transmission of solvent damping into the protein interior is possible. In the Figure 6 results, it should be noted that C_α of Lys-15 is exposed to solvent. In vacuo, long-time (~2 "ps") decay behavior may also occur, as for the Phe-33 C_{ϵ_1} atom in the β sheet; this corresponds to a slow displacement of this residue. The negative parts of the vacuum correlation functions for the C_α and N_ζ atoms in Lys-15 suggest an oscillatory motion with a period of about 3 ps for this residue at the binding site of BPTI.

An estimate of the short-time Cartesian fluctuation relaxation times is obtained by assuming that the correlation function can be written as

$$C(t) = [1 - C(t_M)]e^{-t/\tau} + C(t_M) \quad (1)$$

where $C(t_M)$ has a nonzero value if $C(t)$ has not decayed completely during the period (t_M) available for examination. The values of τ given in Table V are determined with $t = 0.1$ "ps" and $t_M = 2$ "ps"; that is

$$\tau^{-1} = 10 \ln [[1 - C(2.0)]/[C(0.1) - C(2.0)]] \quad (2)$$

For all the atoms listed in the table, the LCS value is longer than the LC value by a factor of 2 or more (except for S_γ of Cys-51). Even for residues with little or no exposed surface area (Gly-36, Asn-43, and Cys-51), there is a significant difference between the solution and vacuum relaxation times; the exposure of each residue to the solvent given in Table V is calculated from the surface area of a residue that is accessible to solvent atoms in the X-ray structure [Deisenhofer & Steigemann, 1977, as obtained from Brookhaven Protein Data Bank, Brookhaven National Laboratory, Upton; Richards (1977) and C. Chothia, private communication]. In both runs, the relaxation times tend to be longer for atoms in residues that are more exposed. The difference in relaxation times observed between the LC and LCS runs is of considerable interest. The most likely origin is the dynamic effect of the solvent viscosity. This might be expected to decrease as a function of the distance of the atom from the protein surface, though if part of the residue (e.g., the OH group of a Tyr side chain) is at or near the surface, transmission of the solvent perturbation to the interior is likely to occur. However, the effects of the solvent on the protein structure could also contribute to the relaxation times. A more detailed examination of the relaxation behavior, including determination of the position dependence of the relaxation times and the simulation of higher viscosity solvents, is in progress.

For the bond angles and improper dihedral angles, the average values and fluctuations are very close to each other in vacuo and in solvent. Because of the large force constants for these degrees of freedom, the presence of the solvent has little effect on their behavior.

Table VI lists the average values and fluctuations for certain dihedral angles. Inclusion of the solvent improves the agreement with the X-ray values, but the overall difference for the ϕ and ψ dihedral angles remains sizable; however, in most cases (particularly for the ϕ values), the dynamic average is in the same minimum (deviation less than ±30°) as the X-ray structure. For the ω angle and the disulfide dihedral angles (C_β–S_γ–S_γ–C_β), the deviations from the X-ray structure are considerably smaller than those for ϕ and ψ. Since the overall structure of the molecule is preserved, differences for neighbor dihedral angles ϕ and ψ compensate approximately, except at the ends of the chain. This is true for the ϕ angles and the ψ angles in different residues, as well as for the well-recognized ψ_{n-1},ϕ_n anticorrelation which tends to confine the structural alterations to oscillations of the peptide group (McCammon et al., 1977).

The rms fluctuations of the dihedral angles in vacuo and in solvent differ very little, not only for the backbone dihedral angles but also when averaged over all dihedral angles in the molecule (Table VI). The fact that the rms fluctuations of the peptide bond angle (ω) are nearly equal to one-half the value for the other dihedral angles (e.g., ϕ and ψ), which have much smaller intrinsic force constants, makes clear the importance of the collisions with the surrounding atoms in determining the range of the fluctuations. The difference between the LCS and LC values for the rms fluctuations of the C_β–S_γ–S_γ–C_β dihedral angles is due to a rearrangement of the Cys-30–Cys-51 disulfide bridge in the LC run (W. F. van Gunsteren and M. Karplus, unpublished results). As for the Cartesian coordinates (Table III), for many of the angles the third moment, $\langle(\Delta q^3)\rangle$, is significantly different from zero, and the fourth moment deviates from the relation $\langle(\Delta q)^4\rangle = 3\langle(\Delta q)^2\rangle^2$ corresponding to harmonic behavior.

The most flexible backbone dihedral angles found during the solvent run occur at the ends of the residue chain and in

Table VI: Dihedral Angle Averages and Fluctuations[a,b]

	Averages						
	$\overline{\langle q \rangle}$			$\overline{q(X)}$,	$\overline{[\langle q \rangle - q(X)]^2}^{1/2}$		
q, dihedrals	LC	LCS	LCC	X-ray	LC/X-ray	LCS/X-ray	LCC/X-ray
C-N-C_α-C (ϕ)					36.0	26.4	39.5
N-C_α-C-N (ψ)					39.4	33.0	41.6
C_α-C-N-C_α (ω)	177.8	178.9	178.9	179.5	10.0	8.8	9.7
C_β-S_γ-S_γ-C_β					5.0	6.6	10.3
all					50.5	42.3	45.1

	Fluctuations					
	$\overline{\langle (\Delta q)^2 \rangle}^{1/2}$			$\overline{\langle (\Delta q)^3 \rangle}^{1/3}$		
q, dihedrals	LC	LCS	LCC	LC	LCS	LCC
C-N-C_α-C (ϕ)	15.7	15.7	15.9	8.6	6.0	2.8
N-C_α-C-N (ψ)	14.5	15.6	15.5	−6.9	3.0	6.6
C_α-C-N-C_α (ω)	7.9	8.0	7.8	−2.2	−0.5	−1.9
C_β-S_γ-S_γ-C_β	10.6	8.5	8.9	−4.0	−4.4	4.4
all	21.3	20.0	25.1	15.6	9.6	−21.8

[a] Average values are listed only for the angle C_α-C-N-C_α (ω) because the other dihedral angles occur in more than one minimum. [b] See footnote *a* of Table III.

Table VII: Dihedral Angle Cross-Correlation Coefficients

		$\frac{\langle \Delta q_1 \Delta q_2 \rangle}{\langle (\Delta q_1)^2 \rangle^{1/2} \langle (\Delta q_2)^2 \rangle^{1/2}}$			exposure
q_1	q_2	LC	LCS	LCC	(%)
Leu-6 ψ	Glu-7 ϕ	−0.86	−0.91	−0.67	53, 25
Arg-17 ψ	Ile-18 ϕ	−0.79	−0.73	−0.76	86, 38
Pro-8 χ^2	Pro-8 χ^3	−0.95	−0.95	−0.96	63
Tyr-10 χ^1	Tyr-10 χ^2	−0.21	0.36	0.28	35
Tyr-21 χ^1	Tyr-21 χ^2	−0.34	−0.28	−0.27	18
Phe-22 χ^1	Phe-22 χ^2	−0.35	−0.38	−0.70	10
Tyr-35 χ^1	Tyr-35 χ^2	−0.46	−0.36	−0.27	6
Arg-39 χ^1	Arg-39 χ^2	0.11	0.46	0.08	79
Arg-39 χ^2	Arg-39 χ^3	0.21	0.07	−0.25	
Arg-39 χ^3	Arg-39 χ^4	0.41	0.59	−0.01	
Arg-39 χ^4	Arg-39 χ^5	−0.07	−0.19	−0.55	

the extended loop (residues 25–28), in agreement with the fact that these regions show the larget atom rms fluctuations. An inspection of the 5-"ps" averages shows that certain dihedral angles jump in their average value by 30–50° after having been rather constant for many picoseconds; e.g., Arg-1 ψ is the same ($\sim$149°) from 0 to 15 ps and then jumps to $\sim$100° where it remains for the rest of the run. The jumps of adjacent dihedral angles are generally correlated (e.g., Arg-1 ψ and Pro-2 ϕ both jump during the 15–20-ps period), but in most cases, the transition does not correspond to a 180° flip of the peptide group (McCammon et al., 1977). In the extended loop, some of the correlations extend over four residues (i.e., that in the 20–25-ps segment). This is indicative of a concerted motion of the loop as a whole. Such longer range correlations differ from the LC run, where only correlations between jumps of adjacent dihedral angles were found.

The equal-time cross-correlation coefficients for a number of neighboring dihedral angles are given in Table VII. As found earlier (McCammon et al., 1977), the backbone ϕ and ψ dihedral angles are generally highly anticorrelated, and the presence of the solvent makes little difference. For the Phe and Tyr side-chain dihedral angles (χ^1 and χ^2), the correlation is smaller, but in most cases it is still negative. This corresponds to a motion dominated by the C_β atom, with the aromatic ring remaining approximately fixed. However, if the ring has considerable space in which to move, the (χ^1 and χ^2) correlation can be positive, as it is for Tyr-10 in the LCS run; this tyrosine ring is on the surface of the protein. For side chains that stick out into the solvent, the correlation coefficients of the LCS and LC runs differ more than for those buried within the protein, as expected. An example is provided by Arg-39 for which there are significant differences in the values of the correlation coefficients, although the signs are the same for the LC and LCS runs.

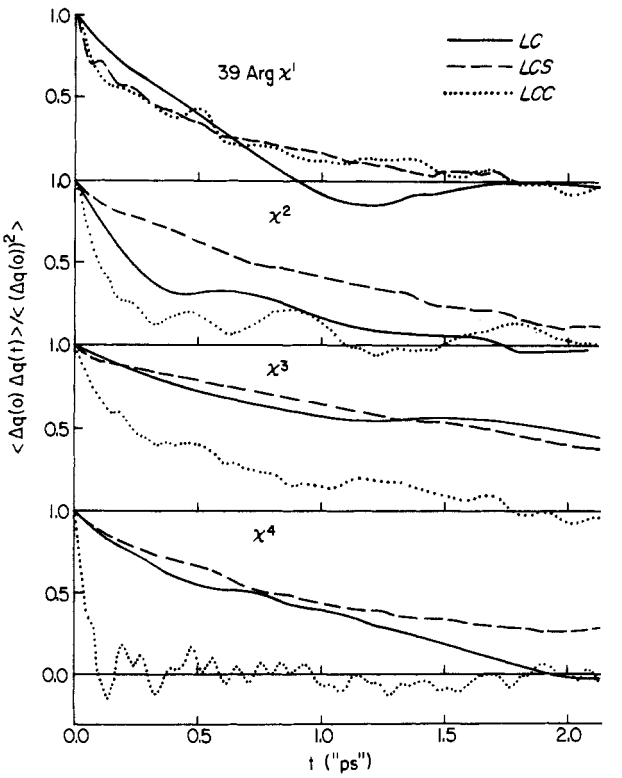

FIGURE 7: Dihedral angle autocorrelation functions for the Arg-39 side chain in the three runs: angles χ^1, χ^2, χ^3, and χ^4.

The dihedral angle autocorrelation functions with and without solvent look very similar, especially for backbone dihedral angles and for dihedral angles of side chains buried in the protein. For side chains exposed to the solvent, small differences are observed. From Figure 7 the LCS autocorrelation functions for the Arg-39 side chain (χ^1–χ^4) are seen to be somewhat less oscillatory than those obtained in vacuo; i.e., the solvent has a damping effect on the dihedral angle motion. The correlation function for χ^5 has a strong oscillatory

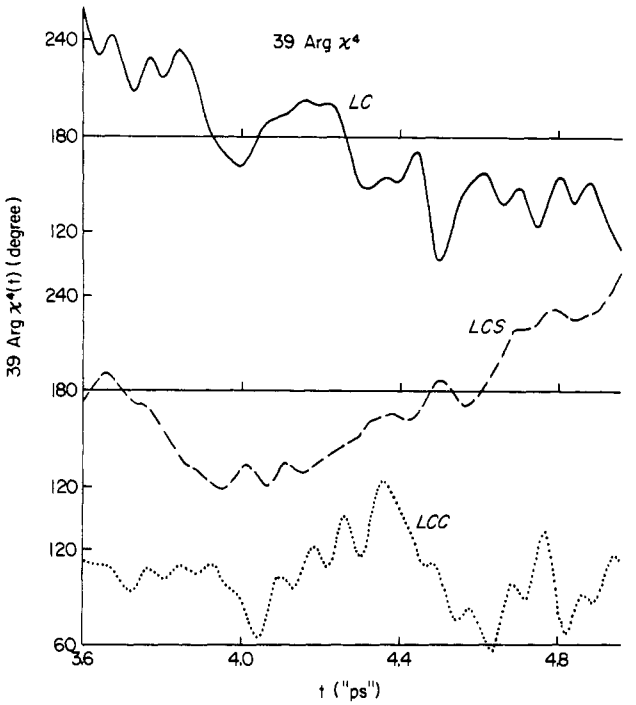

FIGURE 8: A dihedral angle transition: time evolution of the Arg-39 χ^4 angle.

character due to the large torsional force constant associated with this dihedral angle.

The side-chain dihedral angle transitions are strongly affected by solvent. To define a transition, we require that after the dihedral angle has gone through the minimum of a well of the dihedral potential, the angle crosses the minimum of an adjacent well of the potential. With this strict definition, 76 transitions are observed in the LCS run in a period of 25 "ps" (LC: 119), of which 8 occurred in the backbone (LC: 0). They are Ala-25 ψ (60°→−60°), Lys-26 ψ (60°→−60°→60°), Gly-28 ϕ (120°→0°→120°), Gly-36 ψ (−60°→60°→−60°), and Ala-58 ψ (60°→180°). It is seen that the only such transition that is neither in the external loop (25, 26, and 28) nor at the end (58) involves Gly-36, which is at one of the turns of the chain. The presence of backbone dihedral angle transitions in the loop region corresponds to the concerted motion already mentioned. The number of side-chain transitions in solvent is considerably less than in vacuo, since the solvent has a damping effect on the dihedral angle motions. Because the number of side-chain dihedrals that show one or more transitions is not much different in solution (33) and in vacuo (29), the average number of transitions per dihedral angle that shows at least one transition is twice as high in vacuo as in solution. In both runs more than two-thirds of all transitions occur in the charged polar side chains, which tend to stick out into the solvent. It is of interest that the only type of nonpolar residue showing more than four dihedral angle transitions is Leu (χ^2 of Leu-6 and Leu-29). In Figure 8 are displayed transitions (LC and LCS) of Arg-39 χ^4 that happened to occur during the same time period. It takes the dihedral angle typically about 1 ps to wander from one minimum to another minimum of the dihedral potential; the LCS curve shows somewhat more damped behavior. The motion is rather similar to the backbone dihedral angle transition studied in an earlier simulation (McCammon et al., 1977).

To analyze the motion of the aromatic rings in BPTI, we computed the autocorrelation functions

$$\langle P_l(\cos\,[\theta(t)])\rangle = \lim_{\tau\to\infty} \frac{1}{\tau}\int_0^{\tau} P_l[\vec{n}(\tau')\cdot\vec{n}(\tau'+t)]\,d\tau' \quad (3)$$

where

$$\cos\,[\theta(t)] = \vec{n}(0)\cdot\vec{n}(t) \quad (4)$$

and where $l = 1$ and 2 and $\vec{n}$ is the unit vector pointing from the C_{ϵ_1} atom to the C_{ϵ_2} atom in the aromatic ring. These functions, which are of interest for experimental comparisons (P_2 is involved in ^{13}C NMR relaxation and fluorescence depolarization), appear to decay very slowly on the time scale of the simulation. In fact, the P_1 and P_2 correlation functions for all of the rings except Tyr-10 reach a plateau value of $\gtrsim$0.7 in less than 2 "ps" (relaxation time by eq 2 of 1–2 "ps") and then remain essentially constant until the end of the maximum time ($\lesssim$10 "ps") for which valid data are available from the runs. The behavior is to be contrasted with the relaxation time for the time correlation function of the angle, θ, itself, which earlier studies have shown to be in the subpicosecond range (McCammon et al., 1979). For Tyr-10, which is on the surface of the molecule, an approximately exponential decay is found over the period examined, and the estimated relaxation time is 8 "ps". As to the other rings, longer simulations would be required to determine whether there is a significant decay of the P_1 and P_2 angular correlation functions in times shorter than the overall tumbling time of the molecule ($\sim 2 \times 10^{-9}$ s). It is known from NMR data that at room temperature the Tyr rings (other than Tyr-35) undergo 180° rotations, but the rotation times have an experimental upper limit of $\sim 5 \times 10^{-4}$ s (Gelin & Karplus, 1975; Hetzel et al., 1976), much larger than the times of interest here. The solvent seems to have no effect on the rings buried in the protein over the time scale examined in the simulation. In the case of Tyr-10, which is more exposed (35%), the relaxation time in solution appears to be slightly shorter.

To detect larger scale motions of the protein of the "breathing mode" type, we computed the radius of gyration, R_{gyr}, and the particle number density, ρ, as a function of time for various spherical shells around the center of mass of the molecule (see Table VIII); a shell consists of all atoms for which the distance r_{cm} to the center of mass satisfied $r_{c_1} < r_{cm} < r_{c_2}$. The autocorrelation functions of R_{gyr} in vacuo and in solution look very similar. The autocorrelation functions of ρ are displayed in Figure 9 for five different shells. After a rapid initial decay, the LCS curves generally decay somewhat more slowly than those from the LC runs, except for the outermost shell. For this shell the LC curve shows an oscillation with a period of about 2.3 "ps"; the LC curves for the other shells show no such oscillations. In solution (LCS), however, an oscillation with a period of about 1 "ps" can be observed for the fourth shell and less clearly for the other shells, except the first ones. This indicates that BPTI in solution may possess a spherical breathing mode with a period on the order of 1 ps; however, the statistics of the present run are not sufficient to be certain of this result. Treatment of BPTI as a uniform spherical elastic body yields a period of $\sim$1 ps for the lowest frequency breathing mode (Karplus & McCammon, 1981; Suezaki & Go, 1975). The average number density and its rms fluctuation for the different shells are given in the lower part of Table VIII. The large variation in the values for the first shell is due to the fact that this shell contains only 7–9 atoms. For the other shells the fluctuations are very small. The decrease in $\langle\rho\rangle$ for $r_{c_1} > 9$ Å is due to the fact that parts of the shells are outside the protein surface.

The upper part of Table VIII shows that in vacuo the protein has too high a number density and too small a radius of gyration as compared with the X-ray structure. In solution, the calculated values are closer to the X-ray results because of the attractive force exerted by the solvent atoms surrounding the

Table VIII: Radius of Gyration and Number Density[a]

averaging period ("ps")	$\rho(\langle x\rangle)$ (Å^{-3}) (r_{c_1} = 0 Å, r_{c_2} = 8 Å)			$R_{gyr}(\langle x\rangle)$ (Å) (r_{c_1} = 0 Å, r_{c_2} = ∞)		
	LC	LCS (cm)	LCC (cm)	LC	LCS (cm)	LCC (cm)
0-5	0.0695	0.0625	0.0574	10.11	10.68	10.87
5-10	0.0709	0.0625	0.0602	10.11	10.66	10.89
10-15	0.0690	0.0620	0.0597	10.12	10.64	10.85
15-20	0.0634	0.0630	0.0597	10.12	10.62	10.96
20-25	0.0625	0.0639	0.0602	10.11	10.64	10.91
0-25	0.0685	0.0625	0.0597	10.11	10.63	10.89
X-ray	0.0597			10.96		

	LC		LCS		LCC	
$r_{c_1} - r_{c_2}$ (Å)	$\langle\rho\rangle$	$(\langle(\Delta\rho)^2\rangle^{1/2})$ (Å^{-3})	$\langle\rho\rangle$	$(\langle(\Delta\rho)^2\rangle^{1/2})$ (Å^{-3})	$\langle\rho\rangle$	$(\langle(\Delta\rho)^2\rangle^{1/2})$ (Å^{-3})
0-3	0.0704	(0.0105)	0.0582	(0.0100)	0.0827	(0.0089)
3-6	0.0678	(0.0024)	0.0682	(0.0035)	0.0606	(0.0028)
6-9	0.0648	(0.0016)	0.0588	(0.0020)	0.0599	(0.0019)
9-12	0.0340	(0.0011)	0.0325	(0.0012)	0.0320	(0.0010)
12-15	0.0134	(0.0008)	0.0131	(0.0007)	0.0115	(0.0005)

[a] The values of the number density ρ and the radius of gyration R_{gyr} in the upper part of the table are calculated from the various time-averaged structures denoted by $\langle x\rangle$. Atoms that lie between two spherical shells with their origins at the center of mass with radii r_{c_1} and r_{c_2} are included, as indicated.

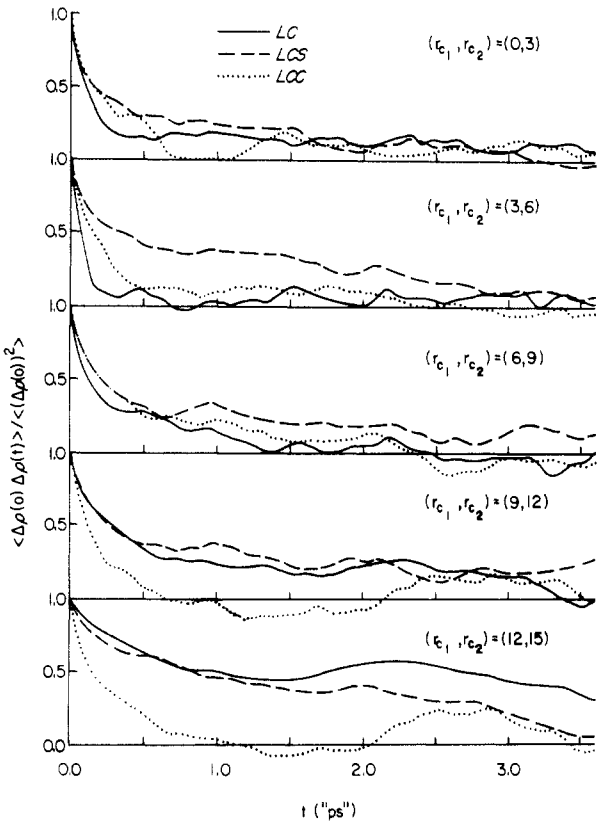

FIGURE 9: Autocorrelation functions of the fluctuation in the number density for concentric shells around the center of mass.

protein. This is one of the effects of the solvent on the average structure of the molecule. However, the deviation from the experimental density in vacuo is sufficiently small that the error in the dynamics is not significant for the magnitude of the fluctuations, as indicated by the very small differences between the LC and LCS runs described in this section.

Summarizing, we conclude that the dynamics of BPTI in a bath of spherical, uncharged solvent atoms with a van der Waals interaction potential are very similar to the dynamics of BPTI in vacuo. The main differences are found in the motion of atoms that are at the surface of the protein. In vacuo, these atoms show a more oscillatory type of motion whereas in solution their motion has a more diffusive character. The root mean square positional fluctuations for these atoms tend to be larger in the presence of solvent, but the actual motions are slower. Further, the side-chain dihedral angles at the surface of the molecule show fewer transitions from one minimum to another in solution than in vacuo. In the interior of the protein, the magnitude of the positional and dihedral angle fluctuations is only slightly altered. There is an indication that the positional relaxation times are increased by the solvent even in the protein interior. Most important, the presence of solvent leads to a time-averaged structure that is considerably closer to the X-ray structure, especially as far as the more flexible parts of the protein are concerned.

The essential role of the solvent seems to be 3-fold. First, it provides an average external force field which prevents the protein from packing too tightly and makes it possible for the parts of the protein that extend outward from the surface not to fold back onto it, as they tend to do in the absence of solvent (loop region, chain ends, and charged polar side chains). The consequence of this solvent effect is an average structure much closer to the X-ray structure than that found in a vacuum simulation. Second, the solvent atoms have a damping effect on the motions of atoms at the protein surface (fewer transitions). The positional relaxation times were increased by factors of 2–3 for many of the protein atoms in the presence of the solvent. Such an effect could be a direct consequence of the increase in the effective external viscosity, though structural changes may be involved as well. The frictional damping effect would be expected to be more important for a solvent with a viscosity closer to that of liquid water. Third, the solvent atoms provide a Brownian driving force which acts on the side chains that stick out into the solvent. For atoms that are buried inside the protein, there seems to be little effect on the dynamics due to the presence of solvent, in accord with the assumption of the earlier molecular dynamics simulations in vacuum (McCammon et al., 1977, 1979; Karplus & McCammon, 1979).

Molecular Dynamics of BPTI in Its Crystalline Environment (LCC)

In this section we compare the molecular dynamics of BPTI in a crystalline environment (LCC) with the vacuum and solvent runs (LC and LCS) to determine the effects of the crystal force field on the equilibrium and dynamic properties of the molecule. Since the force field is taken to be time

independent and is experienced directly only by atoms near the surface of the protein, we do not expect to find very large differences between the LCC results and those obtained in vacuo (LC).

The temperature fluctuations and the heat capacity in the LCC run are the same as those found in vacuo. The velocity autocorrelation functions are similar to those in vacuo (Figure 1), although small differences are observed for atoms that have crystal image neighbor atoms (see Model and Computational Procedure).

The 5-"ps" rms positional fluctuations in crystal and in vacuo are nearly equal (Table II). The significantly smaller value of the all atom rms fluctuations over 25 "ps" in the LCC run vs. the LC run and the similar values found for the 5-"ps" segments suggest that the crystal environment reduces the overall drift of the structure. Comparing the individual atom values in Table II, we see that both the main-chain and side-chain fluctuations tend to be larger in the LC run than in the LCC run. The positional fluctuation autocorrelation functions in general correspond more closely to those in vacuo than to the solvent run (e.g., Lys-15 C_α in Figure 6). This is in accord with relaxation time results in Table V and is what would be expected if the slower LCS relaxation times were the effect of frictional damping due to the solvent viscosity. Certain atoms (e.g., N_ζ of Lys-26 and O_{δ_1} of Asn-43 in Table V) show a large difference between the LCC and LC runs; for Lys-26 the difference originates from the presence of crystal neighbors, but what is happening in the case of Asn-43 is not clear. Also, as shown for Lys-15 N_ζ and Phe-33 C_{ϵ_1} in Figure 6, the correlation function behavior is quite variable; the LCC functions can be closer to the solvent run or differ more from the LC and LCS runs than they do from each other. In some cases, as for the Lys-15 N_ζ atom, a long-time ($\sim$2 ps) oscillation appears to occur.

The rms fluctuations over 5 "ps" of the C_α atoms in the crystal follow roughly the same pattern as in vacuo, except for certain regions of the polypeptide chain (Figure 3); most different is the behavior near the carboxy terminus and of other residues (e.g., Lys-15, Ala-16, and Arg-17) which have crystal neighbors. The large displacements of the atoms around Phe-4 which occur in the 25-"ps" averages are due to the rearrangement of the Cys-5–Cys-55 disulfide bridge in the LCC run at $t \simeq 20$ "ps". The transition takes about 1 ps, just as for other transitions that have been observed (Figure 8; LC and LCS). This suggests that the residues around Phe-4 have two or more separate conformations of low energy in which they can stay for many picoseconds.

The LC and LCC displacements of the side chains displayed in Figure 4 are generally of equal magnitude for those buried inside the protein, but there are differences which can be explained by the environments for side chains at the surface. The LCC rms fluctuations for the Arg-39 and Glu-49 side chains are small because they are hydrogen bonded to crystal image atoms. The Met-52 side chain moves less freely in the crystal than in vacuo due to the presence of the Ala-16 C_β crystal image atom near Met-52 C_ϵ. The motion of the Arg-17 side chain is inhibited by a large number of crystal image neighbors; the Ala-25 O, Lys-26 O, and Ala-58 O_2 atoms bear negative partial charges which attract the positively charged N_ζ atoms in Arg-17 and so tend to keep the side chain in place. All of these results are in accord with the changes in torsional potentials found previously for side chains that interact with crystal neighbors (Gelin & Karplus, 1979).

The distribution of the positional 25-"ps" rms fluctuations of the atoms as a function of their distance, r_{cm}, from the center of mass is different from the LC and LCS results (Figure 5a), in that in the crystal the values for the individual atoms show a larger spread with respect to the average over all atoms than those in vacuo. Further, in the crystal, the distribution does not show as clear a dependence on r_{cm}; however, the backbone atoms do have increased fluctuations as the surface of the molecule is reached. The weak dependence of the LCC results on r_{cm} corresponds to the X-ray temperature factors (Figure 5b); however, the range in the LCC values is much larger than that shown in Figure 5b.

The anisotropy of the atom rms fluctuations is not influenced by the crystal environment (Table III); there is a close correspondence between the values obtained in the LC and LCC runs. Also, the overall results for deviations from harmonic fluctuations are similar. The difference between the 25-"ps" values of C_{ϵ_1} atoms in the crystal and in vacuo is due to the displacement of the Phe-4 and Phe-45 rings in the LCC run caused by the disulfide transition mentioned earlier. For both the anisotropy and the anharmonic character of the fluctuations, the solution results differ more from the LCC and LC runs than they do from each other.

The time-averaged structure, when compared with the X-ray result, shows that the crystal field leads to a considerable improvement relative to in vacuo (Table IV). For the all atom average and for the various atom types, the improvement is comparable to that found when solvent is taken into account (LCS). The parts of the protein whose average structure is most improved by the crystal force field are the extended loop and the carboxy end of the polypeptide chain. The LCC structure is closer to the X-ray structure than to the other structures (LC and LCS). Thus, both LCS and LCC structures are closer to the X-ray structure than they are to each other. This suggests that it is very important to investigate the effect of the simultaneous presence of solvent and a crystal force field.

The rms fluctuations of the backbone dihedral angles in the crystal are very close to those in vacuo and in solvent (Table VI). Averaged over all dihedral angles in the protein, somewhat larger fluctuations are found for the LCC run. This agrees with the slightly larger mobility found for the interior side chains in the crystal and is likely to be due to the lower density of the protein in the crystal (Table VIII). The consecutive 5-"ps" averages of the Cys-51 ϕ and Arg-50 ψ dihedral angles in the LCC run are an example of a dihedral pair that shuttles between two separate low-energy conformations, in each of which it remains for at least a few picoseconds. It is of interest that the differences in the angles of the two minima are rather small; thus, they must arise from the interactions with surrounding protein residues. The transitions from one configuration to the other show the same pattern in crystal and in vacuo. The equal time cross-correlation coefficients for the ϕ and ψ dihedral angles are in general very similar in crystal and in vacuo (Table VII). The positive correlation coefficient for the Tyr-10 χ^1 and χ^2 dihedral angles corresponds to that found for the LCS run and can be explained in the same way [see Molecular Dynamics of BPTI in Solution (LCS)]. The χ^1 and χ^2 dihedral angles in Phe-22 are strongly anticorrelated, because the ring is kept in a fixed position by the crystal image atoms. The Arg-39 side chain is hydrogen bonded to three crystal image atoms, which keep the end of the side chain in a relatively fixed position. This causes the LCC cross-correlation coefficients of the dihedral angles to be negative.

The dihedral autocorrelation functions in the crystal are very similar to those in vacuo, except for side chains at the surface of the protein that experience a strong crystal force field. For the hydrogen-bonded side chain of Arg-39 (Figure 7), the LCC dihedral angle correlation functions are more oscillatory and decay faster the closer they are to the hydrogen-bonded and thus relatively fixed N_ϵ atom (Figure 7).

The number densities and radii of gyration for the three runs are compared in Table VIII. The lower density in the crystal than in vacuo due to the crystal force field has already been described; the same effect was observed for BPTI in the van der Waals solvent. The average strength of the crystal force field in the absence of solvent appears to be approximately correct, since the number density and radius of gyration in the LCC run are very close to the X-ray values. The LCC number density autocorrelation functions are not very different from the ones in vacuo, although for the outermost shells the decay in the crystal is faster than that in vacuo (Figure 9). There appears to be an oscillation with a period of about 2.7 "ps". For the third shell, a weak oscillation with a period of about 1 "ps" can be observed, corresponding to that found in the solvent run.

The results described in this section show that the dynamics of BPTI surrounded by fixed crystal image atoms in the absence of solvent are very similar to the dynamics of BPTI in vacuo. The main overall differences arise from the fact that the molecule is slightly less dense in the crystal than in vacuo and that the average structure is closer to the X-ray structure. Since there is slightly more room for atoms inside the protein to move in the crystal run, the rms fluctuations and the dihedral fluctuations are somewhat larger, and the number of dihedral angle transitions in the crystal increases. For the exterior atoms that strongly interact with crystal image neighbors, the fluctuations are generally smaller and have a higher frequency decay.

Thus, the role of the fixed crystal environment seems to be 2-fold. First, it provides an external force field which prevents the protein from packing too tightly. Second, it holds a number of side chains that stick out into the crystalline environment in relatively fixed positions; the latter is important in keeping the dynamic average structure close to that found in the crystal. Apart from these two specific effects, the presence of the crystal force field has very little effect on the dynamics of the BPTI molecule.

Conclusions

The effects of the surrounding medium on the internal dynamics of proteins are examined theoretically. The results of a molecular dynamic simulation of a protein in solution and of a crystal field environment are presented and compared with those of a simulation in vacuo. For the solution study, the pancreatic trypsin inhibitor and a surrounding medium consisting of 2650 van der Waals solvent atoms at a density equal to that of water were used to construct a periodic system, all of which was included in the simulation. The crystal study was considerably simpler, in that the dynamics of a single protein molecule were simulated in the presence of a static field due to the nearest crystals neighbors. The three simulations (in vacuo, in solution, and in the crystal) were done for an equilibrated system at 300 K with an analysis period of 25 ps.

The effects of a solvent environment can be separated conceptually into changes in the average properties of the protein (e.g., average structure and magnitude of rms fluctuations) and changes in the dynamics (e.g., relaxation times). However, alterations in the average properties can indirectly affect the dynamics. Thus, the solvent can influence the protein dynamics by altering the potential of mean force experienced by the protein atoms, as well as by the frictional (viscosity) and random (Brownian) force terms that act on the protein atoms due to its presence. For determination of the magnitude of these effects, both average and time-dependent properties of the protein calculated from the molecular dynamics simulations in the different environments were compared.

The van der Waals solvent environment was found to have a global effect on the properties of the protein. It provides an external attractive force field that acts on the protein atoms and produces a small, but nonnegligible, reduction in the density of the molecule; the density obtained is very close to the experimental value. Further, the dynamic average structure is found to be significantly closer to the X-ray structure than the structure obtained from the vacuum simulation. The consequences on the dynamics of the structural changes induced by the solvent are small. In the interior of the protein, where the density and other structural effects would be felt, the magnitudes of the positional and dihedral angle fluctuations are very similar in solution and in vacuo. The atoms at the surface of the molecule experience a frictional plus a random driving force due to their direct interaction with the solvent atoms, in addition to the effects of the mean structural changes. The motions of the surface atoms, particularly those in side chains and flexible parts of the molecule, were found to have a more correlated, diffusive character in solution than in vacuo. Further, the root mean square positional fluctuations for these atoms tend to be larger in the presence of solvent. However, the number of dihedral angle transitions is decreased by the solvent damping. In general, the surface atom motions tend to be slowed by the solvent, in that the relaxation times for the positional fluctuation correlation functions are increased. There is also an indication that some interior atom relaxation times are longer in solution, though the details of the behavior have yet to be worked out. The effect of solvent on the relaxation times is made particularly pertinent by the recent experiments of the Frauenfelder group (Beece et al., 1980). They have found that in the rebinding of CO to myoglobin, the external viscosity appears to have a significant effect on the fluctuations of the interior of the protein.

The static crystal environment of the protein was found to have both global and specific effects on the dynamics. The crystal provides an average attractive force field, analogous to that of the solvent. In addition, it binds specifically a number of side chains at the surface of the protein through hydrogen-bond and electrostatic interactions. The dynamic average structure in the crystal field is found to be considerably closer to the X-ray result than the structure obtained from the vacuum run. The magnitudes and fluctuations of most atoms in the protein are little affected by the field. Only for the atoms that have a specific interaction with crystal neighbors are the dynamics altered significantly.

It clearly would be desirable to supplement the present study of environmental effects on protein dynamics by use of a more realistic model of the aqueous solvent. This would increase the frictional drag on the protein atoms due to the higher viscosity, which would be expected to accentuate the diffusional character of the external atom dynamics. Further, the effect of a structured hydrogen-bonding solvent, like the ST2 water model (Stillinger & Rahman, 1972), is likely to be important for polar and charged surface residues. To treat such a problem by a full simulation would require much more computer time, because of both the greater complexity of the model

(Hermans & Rahman, 1976; Hagler & Moult, 1978) and the longer runs needed to obtain reliable statistics on the slower side-chain oscillations, as well as on the more global fluctuations of the protein. Correspondingly, the fixed crystal image atom environment can be improved by including the solvent molecules that occur in the crystal. Finally, it would be desirable to do a complete crystal calculation in which both protein and solvent image atoms are also treated dynamically by performing the appropriate crystal symmetry transformations at every step of the MD run.

Although such extensive calculations are necessary for the complete description of a few systems, the computer time required makes it desirable to search for simpler approaches that introduce the environmental effects. Because the observed differences between the vacuum and solution simulations are rather small, it may be possible to model certain environmental effects without taking explicit account of the individual solvent atoms. The average attractive external force field can be introduced by assuming that the protein is surrounded by a uniform continuous solvent extending to infinity, for which the interaction with the protein can be calculated in a simple fashion. A stochastic dynamic method can be used to model the frictional and random forces exerted on the surface of the protein (Helfand et al., 1970; Levy et al., 1979; Pear & Weiner, 1979). It remains to be investigated whether the inertial terms in the equations of motion may be neglected (Helfand et al., 1970; Levy et al., 1979; Pear & Weiner, 1979) or whether the full Langevin equations have to be integrated to obtain realistic results. In addition, the dielectric effects of the solvent should be introduced. Since the charged side chains yield too strong interactions in the isolated protein, neighboring positive and negative side chains tend to lock each other in relatively fixed positions. Reduction of the values of the charges near the protein surface provides the first approximation to the dielectric shielding (Northrup et al., 1980a,b).

In the molecular dynamics calculations, there are observed certain structural transitions that differ in the various runs. These are likely to represent rare events that take place by accident in a given simulation. One type of transition that was found involves the disulfide bonds; both the 30–51 bond (LC run) and the 5–55 bond (LCC run) undergo a transition and remain in the new structure during a significant part of the trajectory. Thus, it appears that two metastable structures that are similar, but not identical, exist in the neighborhood of the S–S bonds.

In summary, the present study demonstrates that the most important influence of a solvent or crystal environment is on the dynamic average structure of the protein; it is considerably closer to the X-ray structure than that obtained from a vacuum simulation. The magnitudes of the fluctuations of atoms in the interior of the protein are little influenced by the environment. However, there is an indication that there is an increase in the relaxation times of these fluctuations. For atoms of side chains at the surface of the protein, the solvent or crystal environment alters both the magnitudes and time course of the fluctuations. The changes are such that it may be possible to model the solvent effect by average force fields and by stochastic dynamics methods.

Acknowledgments

We thank T. Ichiye, R. M. Levy, B. Olafson, and S. Swaminathan for helpful discussions. M. Karplus thanks J.-M. Lehn and the Laboratoire de Chimie des Interactions Molēculaires, Collège de France, Paris, France, for hospitality during the time that the final draft of the manuscript was completed.

References

Artymuik, P. J., Blake, C. C. F., Grace, D. E. P., Oatley, S. J., Phillips, D. C., & Sternberg, M. J. E. (1979) *Nature (London) 280*, 563.

Beece, D., Eisenstein, L., Frauenfelder, H., Good, D., Marden, M. C., Reinisch, L., Ryenolds, A. H., & Sorensen, L. B. (1980) *Biochemistry 19*, 5147.

Bishop, M., Kalos, M. H., & Frisch, H. L. (1979) *J. Chem. Phys. 70*, 1299.

Campbell, I. D. (1977) in *NMR in Biology* (Dwek, R. A., et al., Eds.) Academic Press, New York.

Chandler, D. (1974) *Acc. Chem. Res. 7*, 246.

Deisenhofer, J. O., & Steigemann, W. R. (1977) *Acta Crystallogr., Sect. B B31*, 238.

Frauenfelder, H., Petsko, G. A., & Tsernoglou, D. (1979) *Nature (London) 280*, 558.

Geiger, A., Rahman, A., & Stillinger, F. H. (1979) *J. Chem. Phys. 70*, 263.

Gelin, B. R., & Karplus, M. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2002.

Gelin, B. R., & Karplus, M. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 801.

Gelin, B. R., & Karplus, M. (1979) *Biochemistry 18*, 1256.

Hagler, A. T., & Moult, J. (1978) *Nature (London) 272*, 222.

Helfand, E., Wasserman, Z. R., & Weber, T. A. (1970) *J. Chem. Phys. 70*, 2016.

Henry, N. F. M., & Lonsdale, E. (Eds.) (1965) *International Tables for X-ray Crystallography*, Vol. I, Kynoch Press, Birmingham.

Hermans, J., & Rahman, A. (1976) in *Models for Protein Dynamics* (Berendsen, H. J. C., Ed.) CECAM, Orsay.

Hetzel, R., Wüthrich, K., Deisenhofer, J., & Huber, R. (1976) *Biophys. Struct. Mech. 2*, 159.

Hockney, R. W., Goel, S. P., & Eastwood, J. W. (1974) *J. Comput. Phys. 14*, 148.

Huber, R., Kukla, D., Rühlmann, A., Epp, O., & Formanek, H. (1970) *Naturwissenschaften 57*, 389.

Karplus, M., & McCammon, J. A. (1979) *Nature (London) 277*, 578.

Karplus, M., & McCammon, J. A. (1981) *CRC Crit. Rev. Biochem. 9*, 283.

Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1.

Lakowicz, J. R., & Weber, G. (1973) *Biochemistry 12*, 4171.

Levy, R. M., & Karplus, M. (1979) *Biopolymers 18*, 2465.

Levy, R. M., Karplus, M., & McCamman, J. A. (1979) *Chem. Phys. Lett. 65*, 49.

Matthews, B. W. (1976) *Annu. Rev. Phys. Chem. 27*, 493.

McCammon, J. A., & Karplus, M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3585.

McCammon, J. A., Gelin, B. R., & Karplus, M. (1977) *Nature (London) 267*, 585.

McCammon, J. A., Wolynes, P. G., & Karplus, M. (1979) *Biochemistry 18*, 927.

McLachlan, A. D. (1979) *J. Mol. Biol. 128*, 49.

McQuarrie, D. A. (1973) *Statistical Mechanics*, Chapter 13, Harper and Row, New York.

Northrup, S. H., Pear, M. R., McCammon, J. A., Karplus, M., & Takano, T. (1980a) *Nature (London) 287*, 659.

Northrup, S. H., Pear, M. R., McCammon, J. A., & Karplus, M. (1980b) *Nature (London) 286*, 304.

Owicki, J. C., & Scheraga, H. A. (1977) *J. Am. Chem. Soc. 99*, 7413.

Pangali, C., Rao, M., & Berne, B. J. (1979) *J. Chem. Phys. 71*, 2975, 1299.

Pear, M. R., & Weiner, J. H. (1979) *J. Chem. Phys. 71*, 212.
Quentrec, N., & Brot, C. (1973) *J. Comput. Phys. 13*, 430.
Rabolt, J. F., Moore, W. H., & Krimm, S. (1977) *Macromolecules 10*, 1065.
Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151.
Rigler, R. (1977) *Trends Biochem. Sci. 2*, 252.
Rossky, P. J., & Karplus, M. (1979) *J. Am. Chem. Soc. 101*, 1913.
Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327.
Saviotti, M. L., & Galley, W. C. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 4154.
Shepherd, I. W. (1975) *Rep. Prog. Phys. 38*, 565.
Shimanouchi, T. (1976) *Structural Studies of Macromolecules by Spectroscopic Methods* (Ivin, K. J., Ed.) Wiley, New York.
Stillinger, F. H., & Rahman, A. (1972) *J. Chem. Phys. 57*, 1281.
Suezaki, Y., & Go, N. (1975) *Int. J. Pept. Protein Res. 7*, 333.
van Gunsteren, W. F. (1980) *Mol. Phys. 40*, 1015.
van Gunsteren, W. F., & Berendsen, H. J. C. (1977) *Mol. Phys. 34*, 1311.
van Gunsteren, W. F., & Karplus, M. (1980) *J. Comput. Chem. 1*, 266.
Verlet, L. (1967) *Phys. Rev. 159*, 98.
Wagner, G., & Wüthrich, K. (1978) *Nature (London) 275*, 247.
Weber, T. A., & Helfand, E. (1979) *J. Chem. Phys. 71*, 4760.
Woodward, C. K., & Hilton, B. D. (1979) *Annu. Rev. Biophys. Bioeng. 8*, 99.
Yu, N.-T. (1977) *CRC Crit. Rev. Biochem. 4*, 229.

Thermodynamics and Kinetics of Single Residue Replacements in Avian Ovomucoid Third Domains: Effect on Inhibitor Interactions with Serine Proteinases†

Mark W. Empie and Michael Laskowski, Jr.*

ABSTRACT: Sequence determinations in our laboratory have yielded the primary structures of ovomucoid third domains from 35 avian species. From this list, 12 sequences could be arranged into a contiguous set such that each sequence differs from a second by a single amino acid replacement. For this set of domains and for five additional domains of special interest, we report here the association equilibrium constants for their binding with bovine α-chymotrypsin, elastase I, and subtilisin Carlsberg. The results are interpreted with the aid of the three-dimensional structure of highly homologous Japanese quail ovomucoid third domain and of computer-generated models of the complexes of the inhibitor with the respective enzymes. The results show that (i) changes in inhibitor residues other than the primary recognition residue (P_1), even sequentially far from the reactive site, may exert large effects on association equilibrium constant values provided these residues make contact with the enzyme, (ii) changes in residues other than P_1 often exert large differential effects toward the different enzymes, i.e., the same change can make the inhibitor stronger for one enzyme and weaker for another, (iii) the sign and to some extent the magnitude of the changes can be rationalized from the known structures of the inhibitor and the enzyme, (iv) changes in surface residues which do not contact the enzyme in complex are virtually without effect, and (v) glycosylated and nonglycosylated inhibitors have the same constants. For confirmation of the validity of the equilibrium constant comparisons in a few cases, the rate constants k_{on} and k_d were determined and the resultant calculated equilibrium constant values compared to the directly determined numbers. An additional test of validity is provided by experiments where a glycosylated domain of one species is allowed to compete with an unglycosylated domain of another for the same enzyme.

It is a major objective of this laboratory to construct an algorithm which would predict, from only the amino acid sequence, the thermodynamic and kinetic parameters for the interaction of serine proteinases with their protein inhibitors (Laskowski, 1980; Laskowski et al., 1981). The kinetic and thermodynamic properties are known for a number of these protein inhibitors which follow an analogous, standard mechanism (Laskowski & Sealock, 1971; Laskowski & Kato, 1980). However, not all these inhibitors are homologous; instead, they can be grouped into several families (Laskowski & Kato, 1980). In development of the algorithm, it is essential to focus first on a single family. To date, systematic studies within a family have been limited and have focused on the two residues, P_1 and P_1' (Schecter & Berger, 1967), which make up the reactive site peptide bond (Sealock & Laskowski, 1969; Kowalski et al., 1974; Kowalski & Laskowski, 1976; Odani & Ikenaka, 1978; Odani & Ono, 1980; Wenzel & Tschesche, 1981). For a variety of reasons, but mainly the availability of needed, naturally occurring variants, we have chosen to focus on the Kazal family of inhibitors. In particular, this paper concentrates on a subset of this family—the third domain of avian ovomucoid.

Ovomucoid is a major component (~10% of the protein) of avian egg whites and is responsible for most of the inhibitory activity against serine proteinases in the egg white. Intact ovomucoid was found by Feeney and co-workers (Rhodes et al., 1960; Feeney, 1971) to have two startling characteristics: (1) a single ovomucoid molecule, depending on the avian

† From the Department of Chemistry, Purdue University, West Lafayette, Indiana 47907. *Received November 13, 1981.* Supported by Grant GM 10831 from the National Institutes of General Medical Sciences, National Institutes of Health. Many of the results given here were presented in preliminary form (Empie & Laskowski, 1981; Laskowski et al., 1981).

0006-2960/82/0421-2274$01.25/0