Effect of Constraints on the Dynamics of Macromolecules†

Wilfred F. van Gunsteren‡ and Martin Karplus*

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138.
Received December 16, 1981

ABSTRACT: The effects of constraints on the dynamics of macromolecules is investigated by the method of computer simulation. Three molecular dynamics simulations of a protein, bovine pancreatic trypsin inhibitor (BPTI), consisting of 454 heavy atoms, are compared; one is made without applying any constraints, one with bond-length constraints, and one with bond-length and bond-angle constraints. Equilibrium averages and time-dependent molecular properties are examined to determine the effects on the dynamics of the macromolecule of constraining the bond-length and bond-angle degrees of freedom. The use of fixed covalent bond lengths does not significantly alter the dynamical properties of the protein on a time scale longer than 0.05 ps. This makes it possible to obtain a threefold increase of the computational efficiency of macromolecular simulations by the application of bond-length constraints. Constraint of the bond angles has an important effect on the dynamics. The magnitude of the fluctuations (root-mean-square Cartesian and internal coordinate fluctuations) is decreased by a factor of 2 and the dihedral transition rates are dramatically reduced. This makes clear that in a closely packed system, such as a protein in its native configuration, the excluded volume effects due to repulsive van der Waals interactions introduce a strong coupling between the dihedral angle and bond-angle degrees of freedom. A detailed model of the equilibrium or dynamic properties of such systems must therefore take account of the role of bond-angle fluctuations.

1. Introduction

Computer simulation techniques have been shown to be a powerful tool for probing the properties of liquids or solutions of flexible molecules. Molecular dynamics (MD), in which Newton's equations of motion are solved for an equilibrated system, has been used to model the dynamics of *n*-alkanes,[1-7] of a dipeptide dissolved in water,[8,9] and of proteins.[10-14] The Monte Carlo method (MC) has been used in combination with MD to model the trajectories of activated processes in proteins.[14] Recently, the dynamics of short polymers have been simulated by using stochastic dynamics (SD), in which the Langevin or related equations of motion for the system are solved.[15-21]

For most of the flexible molecules that have been studied, the covalently bonded structure confines some of the internal coordinates to a narrow range of values at ordinary temperatures. Since the bond-stretching frequencies are of the order of 30 ps^{-1} and the bond-angle bending frequencies are of the order of 10 ps^{-1}, root-mean-square (rms) fluctuations of the order of 0.03 Å for bond lengths and 5° for bond angles are expected in the harmonic approximation for molecules composed of first-row atoms. It has often been assumed, therefore, in statistical mechanical and dynamical descriptions of flexible molecules (especially macromolecules) that the bond lengths, the bond angles, and even some of the dihedral angles (e.g., peptide bond dihedral angles) can be treated as rigidly constrained.[22,23] In this way the highest frequency components of the molecular motion are eliminated and the size of the time step used in integrating the equations of motion (MD, SD) can be increased. Thus, the computational efficiency of MD or SD simulations can be improved by applying constraints to some of the degrees of freedom. A similar argument holds for MC or energy minimization techniques[24] with constraints. Because of the great promise of simulation methods for providing structural and dynamic as well as thermodynamic information about complex systems, it is essential to determine whether the flexible molecules of interest are adequately modeled in the presence of bond-length or bond-angle constraints. The best approach to this problem, particularly for large systems where analytic studies are difficult, is a comparison of the results obtained for the properties of interest from simulations with and without constraints.

It is the purpose of this paper to evaluate the effect of constraints on the molecular dynamics of a globular protein, that is, to determine whether the introduction of bond-length and bond-angle constraints in simulations of such densely packed systems yields the same statistical properties as do unconstrained dynamic trajectories. To introduce the necessary constraints, we use the SHAKE method,[25,26] which is especially appropriate for macromolecules since it treats the constraints one after another in an iterative way; the increase in the required computer time thus is proportional to the number of atoms N. The molecule used for investigating the effect of constraints is a protein, bovine pancreatic trypsin inhibitor (BPTI). This molecule was chosen for study because there have been a number of previous simulations of its dynamic properties.[10-12] Three molecular dynamics runs have been performed: one without any constraints (NC), one with bond-length constraints (LC), and one with bond-length and bond-angle constraints (LAC). To compare the three MD runs and determine whether or not they sample approximately the same part of phase space, a variety of statistical properties are analyzed; they include the averages, fluctuations, and correlation functions for various physical quantities.

The protein model is described in section 2. The computational procedure of the MD runs is outlined in section 3. In section 4 the simulation without constraints is described and compared with earlier simulation studies of BPTI.[10-12] The runs with bond-length constraints and with bond-length and bond-angle constraints are analyzed in sections 5 and 6. Section 7 contains the conclusions.

2. Protein Model

The protein model and computer program are very similar to those employed in the previous simulations[10-12,27] but differ from them in a number of details, which do not affect the present analysis or comparisons with earlier work. However, to make clear the basis of the present calculation, a complete specification of the model is given. BPTI consists of 58 amino acid residues. It is represented

† Supported in part by grants from the National Science Foundation, the National Institutes of Health, and The Netherlands Organization for the Advancement of Pure Research (ZWO).

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

0024-9297/82/2215-1528$01.25/0

Table I
Atom Types, Masses, and van der Waals Parameters[a]

atom type	atom symbol	group represented	mass, u	$\alpha/4\pi\epsilon_0$, Å^3	N	R, Å	occurring in
1	O	carbonyl oxygen	15.9994	0.84	6	1.60	all residues
2	OH1	alcoholic hydroxyl	17.0074	1.20	7	1.65	Asph, Gluh, Hyp, Ser, Thr, Tyr, ct
3	OM	carboxyl oxygen	15.9994	2.14	6	1.60	Asp, Glu, Heme, ct
4	NH1	NH group	15.0147	1.40	7	1.65	all residues but Hyp, Pro; at
5	NH2	NH_2 terminal	16.0227	1.70	8	1.70	Asn, Gln, Lys, at
6	NH3	NH_3 terminal	17.0307	2.13	9	1.75	Lysh, at
7	CH1	aliphatic CH group	13.019	1.35	6	1.85	all residues but Gly
8	CH2	aliphatic CH_2 group	14.027	1.77	7	1.90	all residues but Ala, Thr, Val
9	CH3	CH_3 terminal	15.035	2.17	8	1.95	Ala, Ile, Leu, Met, Thr, Val, Heme, mt
10	C	bare carbon	12.011	1.65	5	1.80	all residues, Heme
11	CR1	aromatic CH group	13.019	2.07	6	1.90	His, Phe, Trp, Tyr, Heme
12	S	sulfur	32.06	0.34	16	1.90	Cys, Met
13	N	bare peptide nitrogen	14.0067	1.10	6	1.60	Hyp, Pro
14	NP	porphyrin nitrogen	14.0067	1.10	6	1.60	Heme
15	FE	heme iron	55.847	0.01	0.0001	0.65	Heme
16	OH2	solvent H_2O	18.0154	1.20	7	1.70	solvent
17	SH1	SH group	33.068	0.34	16	1.95	Cysh
18	NR	aromatic nitrogen	14.0067	1.10	6	1.60	Hisa, Hisb
19	NC2	charged NH_2 terminal	16.0227	1.70	8	1.70	Arg

[a] The symbol α denotes the polarizability, N the number of effective outer-shell electrons, R the van der Waals radius to be used in the Slater-Kirkwood formula (eq 2). Some residues can have differently charged side chains. They are distinguished by adding the character h to the residue name for the most positively charged configuration. The assignment "all residues" means Ala, Arg, Asn, Asp, Asph, Lys, Lysh, Gln, Glu, Gluh, Gly, Hisa, Hisb, Hyp, Ile, Leu, Lys, Lysh, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val. Histidine having N_{δ_1} protonated but not N_{ϵ_2} is denoted by Hisa; when N_{ϵ_2} is protonated but not N_{δ_1}, it is denoted by Hisb; Hish is doubly protonated; when the protonation state plays no role the symbol His is used. Various atom types or groups of atom types may occur at the amino end of the chain. They are denoted by the symbols at (NH, NH_2^+, NH_3^{2+}), mt (CH_3), or act (CH_3CO). At the carboxy end of the chain they are denoted by ct (COOH, COO^-) or mt (CH_3NH).

as a set of extended atoms; that is, hydrogen atoms are not explicitly considered but incorporated into the heavy atoms to which they are bound. BPTI consists of 454 extended atoms; in addition, four internally hydrogen-bonded water molecules are included in the simulations, making the total number of atoms equal to $N_{at} = 458$. The set of (extended) atom types[12,27] representing the various atoms or atom groups occurring in the residues most commonly found in proteins is given in Table I. Although the residues His, Hyp, and Trp do not occur in BPTI, their model parameters are included in the description below for reasons of completeness. The parameters for the heme group, such as would be required by hemoglobin,[27] are not listed, but when amino acid atom types of parameters could also be used for a heme group, they are denoted by the symbol Heme. The pH value is assumed to be between 5 and 9.5; that is, the amino terminus and the carboxy terminus of the protein consist of atoms of type NH3 (NH_3^+) and OM(COO^-), respectively; correspondingly, the acidic and basic amino acids of the protein are assumed to be charged.

The energy function is composed of terms representing covalent bond stretching, bond-angle bending, harmonic (out-of-plane, out-of-tetrahedal configuration) dihedral bending, sinusoidal dihedral torsion, and van der Waals, electrostatic (Coulomb), and hydrogen-bond interactions; that is, the energy function has the form

$$V(\vec{r}) \equiv V(\vec{r}_1,\vec{r}_2,...,\vec{r}_{N_{at}}) = \frac{1}{2}\sum_{l=1}^{N_b} K_{b_l}[b_l - b_{0_l}]^2 + \frac{1}{2}\sum_{l=1}^{N_\theta} K_{\theta_l}[\theta_l - \theta_{0_l}]^2 + \frac{1}{2}\sum_{l=1}^{N_\xi} K_{\xi_l}[\xi_l - \xi_{0_l}]^2 + \sum_{l=1}^{N_\phi} K_{\phi_l}[1 + \cos(n_l\phi_l - \delta_l)] + \sum_{i<j}^{N_{at}} [C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^6 + q_iq_j/(4\pi\epsilon_0\epsilon_r r_{ij})]S(r_{ij}) + \sum_{(i,j)=1}^{N_{hb}} [C_{12}'(i,j)/r_{ij}^{12} - C_{10}'(i,j)/r_{ij}^{10}] \quad (1)$$

The expression for $V(\vec{r})$ is similar to the empirical potentials of ref 10–12 and 27–33. It is noted that the dihedral angles in eq 1 are divided into two classes: dihedral angles ξ that keep certain atom groups near planar or tetrahedral configurations (they obey a harmonic potential and are called "improper" or harmonic dihedral angles) and dihedral angles ϕ that can perform complete (360°) rotations (they obey a sinusoidal potential and are called simply dihedral angles). In eq 1 the Cartesian position vectors of the atoms are denoted by $\vec{r}_1, \vec{r}_2, ..., \vec{r}_{N_{at}}$. In the first term, the summation runs over all N_b covalent bonds l occurring in the protein; for BPTI, $N_b = 468$. The current bond length of bond l is denoted by b_l and the values of the interaction function parameters K_{b_l} and b_{0_l} are given in Table II as a function of atom type pairs. The summation in the second term runs over all N_θ bond angles l; for BPTI, $N_\theta = 630$. The current bond angle is denoted by θ_l and the interaction function parameters K_{θ_l} and θ_{0_l} are given in Table III. An improper dihedral angle, A–X–Y–B, is defined as the angle between the plane through atoms A, X, and Y and the plane through the atoms X, Y, and B. In Table IV it is specified which atoms in the various residues form improper dihedrals. The interaction parameters have been chosen as follows. The improper dihedrals corresponding to a tetrahedral configuration (i.e., those for which the first or fourth atom is of atom type CH1) have $\xi_0 = 35.26439°$ and $K_\xi = 80$ kcal mol^{-1} rad^{-2}. All others have $\xi_0 = 0°$ and $K_\xi = 40$ kcal mol^{-1} rad^{-2}. We note that the IUPAC–IUB convention[34] on atom names and dihedral angles is used throughout this paper; that is, the cis conformation has $\phi = 0°$ or $\xi = 0°$. In Table V the atoms that form dihedrals ϕ in the various residues are specified. The dihedral interaction parameters have been chosen according to the atom type of the two central atoms forming the dihedral angle. They are given in Table VI; for BPTI $N_\xi = 196$ and $N_\phi = 300$.

The van der Waals parameters for an atom pair (i,j) are

Table II
Bond Parameters[a]

bond type	$^1/_2K_b$, kcal mol^{-1} Å^{-2}	b_0, Å	occurring in
CH1-OH1	400	1.42	Hyp, Thr
CH1-NH1	450	1.46	all residues but Gly, Hyp, Pro; at
CH1-NH2	450	1.47	at
CH1-NH3	450	1.47	at
CH1-CH1	400	1.53	Ile, Thr, Val
CH2-OH1	400	1.42	Ser
CH2-NH1	450	1.46	Arg, Gly, at
CH2-NH2	450	1.48	Lys, at
CH2-NH3	450	1.48	Lysh, at
CH2-CH1	400	1.52	all residues but Ala, Gly, Thr, Val
CH2-CH2	400	1.52	Arg, Gln, Glu, Gluh, Lys, Lysh, Met, Pro, Heme
CH3-NH1	450	1.47	mt
CH3-CH1	400	1.52	Ala, Ile, Leu, Thr, Val
CH3-CH2	400	1.54	Ile
C-O	600	1.23	all residues
C-OH1	450	1.38	Asph, Gluh, Tyr, ct
C-OM	450	1.22	Asp, Glu, Heme, ct
C-NH1	500	1.32	all residues but Hyp, Pro
C-NH2	450	1.33	Asn, Gln
C-CH1	400	1.52	all residues but Gly
C-CH2	400	1.52	Asn, Asp, Asph, Gln, Glu, Gluh, Gly, His, Phe, Trp, Tyr, Heme
C-CH3	400	1.52	Heme, act
C-C	500	1.39	Trp, Heme
CR1-NH1	500	1.32	His, Trp
CR1-C	500	1.39	His, Phe, Trp, Tyr, Heme
CR1-CR1	500	1.39	Phe, Trp, Tyr
S-CH2	450	1.81	Cys, Met
S-CH3	450	1.77	Met
S-S	500	2.02	Cys-Cys
N-CH1	450	1.46	Hyp, Pro
N-CH2	450	1.46	Hyp, Pro
N-CH3	450	1.47	mt
N-C	500	1.32	Hyp, Pro
SH1-CH2	450	1.81	Lysh
NR-C	500	1.32	Hisb
NR-CR1	500	1.32	Hisa, Hisb
NC2-C	450	1.33	Arg

[a] See footnote of Table I.

obtained from the Slater–Kirkwood formula[35]

$$C_6(i,j) = (4\pi\epsilon_0)^{-2}(^3/_2 e\hbar m_e^{-1/2})\alpha_i\alpha_j[(\alpha_i/N_i)^{1/2} + (\alpha_j/N_j)^{1/2}]^{-1} \quad (2)$$

where ϵ_0 is the permittivity of vacuum, e is the electron charge, $\hbar$ is Planck's constant, and m_e is the electron rest mass, and $C_{12}(i,j)$ is obtained from

$$C_{12}(i,j) = {}^1/_2 C_6(i,j)[R_i + R_j]^6 \quad (3)$$

The polarizability α_i, the number of effective outer-shell electrons N_i, and the van der Waals radius R_i are given in Table I for various atom types. The constant in eq 2 equals 362.3461 kcal mol^{-1} Å$^{3/2}$.[36] All atoms have partial charges. They are listed in ref 12. We note that the charge of the O atom in Arg has been misprinted in Table AIII of ref 12; it is −0.360 e. The dielectric permittivity is assumed to equal $\epsilon_r \equiv r_{ij}$ Å^{-1}, as in ref 12. The constant $(4\pi\epsilon_0)^{-1}$ has the value 332.0716 kcal mol^{-1} e^{-2} Å. The summation of the nonbonded term runs only over pairs of atoms that are separated by three or more covalent bonds. For BPTI this number of pairs is equal to about 100 000. It can be considerably reduced by applying a cutoff radius R_c beyond which no interactions are included. To avoid a discontinuity in the potential and the force at the cutoff radius, a switching function $S(r)$ is applied. We use

$$S(r) \equiv 1 \qquad r \le R_s \quad (4a)$$

$$S(r) \equiv [R_c^2 - r^2]^2[R_c^2 + 2r^2 - 3R_s^2][R_c^2 - R_s^2]^{-3} \qquad R_s \le r \le R_c \quad (4b)$$

$$S(r) \equiv 0 \qquad r \ge R_c \quad (4c)$$

which has the required continuity properties and is a function of r^2, which is computationally covenient. The parameters in eq 4 were chosen to be R_s = 7.5 Å and R_c = 8.0 Å; this reduces the number of atom pairs to about 20 000.

The last term in eq 1 models the hydrogen-bond interaction function. The parameters for a donor–acceptor pair (i,j) can be obtained from Table VII, with

$$C_{12}'(i,j) = -5E_{min}(i,j)R_{min}^{12}(i,j) \quad (5)$$

and

$$C_{10}'(i,j) = -6E_{min}(i,j)R_{min}^{10}(i,j) \quad (6)$$

A donor–acceptor pair (i,j) is considered as a hydrogen-bonded pair when the following criteria are satisfied:

1. $r_{ij} \le 3.30$ Å (7)
2. $|n_i - n_j| \ge 3$ (8)

where n_i denotes the residue number of i

3. $r_{hj} < 2.5$ Å (9)

where the hydrogen atom bound to the donor is denoted by h

4. $\theta(i\text{–}h\text{–}j) > 150°$ (10)
5. $\theta(h\text{–}j\text{–}k) > 80°$ (11)
6. $|\phi(h_1\text{–}j\text{–}k\text{–}l) - \phi(h_2\text{–}j\text{–}h\text{–}l)| > 120°$ (12)

where k denotes the neighbor atom of the acceptor j, l is a neighbor of atom k, and hydrogens competing for the same acceptor are denoted by h_1 and h_2.

The X-ray coordinates[37] are used in selecting hydrogen bonds, and the hydrogen coordinates are assumed to satisfy standard geometries. The donor–hydrogen distance is r_{ih} = 1 Å. In OH1, the angle $\theta(x\text{–}i\text{–}h) = 120°$, where the neighbor of the donor is denoted by x. If free rotation of the hydrogen around the x–i axis is possible, the criteria are checked for the different hydrogen positions. In OH2, the angle $\theta(h_1\text{–}i\text{–}h_2)$ is tetrahedal. In NH1, the hydrogen lies in the plane through atoms x, y, and i (x and y are neighbors of i), with $\theta(x\text{–}i\text{–}h) = 120°$. In NH2 and NC2, we have $\theta(x\text{–}i\text{–}h_1) = \theta(x\text{–}i\text{–}h_2) = \theta(h_1\text{–}i\text{–}h_2) = 120°$. In Arg, the two planar NC2 configurations lie in one plane. In NH3, the i–x, i–h_1, i–h_2, and i–h_3 directions are tetrahedal. For BPTI, N_{hb} = 45 hydrogen-bonded pairs are found, which have been listed in Table AIV of ref 12. The atom names used in ref 27 differ from those used in ref 37. The N_{η_1} and N_{η_2} atoms have been interchanged in all Arg residues and the first and second internal water molecules have been interchanged with the third and fourth, respectively; here, the assignment of ref 37 is followed. We note that hydrogen-bonded atom pairs interact through the last two terms in eq 1.

When bond-length constraints are applied, the first term in eq 1 is omitted and all bond lengths of the protein are kept fixed. When bond angles are constrained, too, the first three terms in eq 1 are omitted, but not all bond angles of the protein can be kept fixed, since they would overdetermine the constraints. The bond angles that are not constrained are listed in Table VIII. For BPTI, 622 out of the 630 bond angles form the set of bond-angle

Table III
Bond-Angle Parameters[a]

bond-angle type	$^1/_2K_\theta$, kcal mol^{-1} rad^{-2}	θ_0, deg	occurring in
CH3-NH1-CH1	70	117.4	mt
CH3-NH1-CH2	70	117.4	mt
C-NH1-CH1	60	119.1	all residues but Gly
C-NH1-CH2	60	117.8	Arg, Gly
C-NH1-CH3	60	117.8	mt
CR1-NH1-C	50	108.5	Hisa, Hish, Trp
CR1-NH1-CR1	50	109.6	Hisb, Hish
CH2-NH2-CH1	70	117.4	at
CH1-CH1-OH1	50	101.5	Thr
CH1-CH1-NH1	50	107.1	Ile, Thr, Val
CH1-CH1-NH2	50	108.9	at
CH1-CH1-NH3	50	108.9	at
CH2-CH1-OH1	30	110.5	Hyp
CH2-CH1-NH1	30	109.0	all residues but Ala, Gly, Hyp, Ile, Pro, Thr, Val
CH2-CH1-NH2	30	109.8	at
CH2-CH1-NH3	30	109.8	at
CH2-CH1-CH1	30	108.9	Ile
CH2-CH1-CH2	30	108.9	Hyp
CH3-CH1-OH1	30	110.5	Thr
CH3-CH1-NH1	30	108.6	Ala
CH3-CH1-NH2	50	108.9	at
CH3-CH1-NH3	50	108.9	at
CH3-CH1-CH1	50	107.8	Thr, Val
CH3-CH1-CH2	30	108.9	Ile, Leu
CH3-CH1-CH3	30	110.7	Leu, Val
C-CH1-NH1	60	108.6	all residues but Gly, Hyp, Pro
C-CH1-NH2	30	108.0	at
C-CH1-NH3	30	108.0	at
C-CH1-CH1	30	107.5	Ile, Thr, Val
C-CH1-CH2	30	107.1	all residues but Ala, Gly, Ile, Thr, Val
C-CH1-CH3	50	104.3	Ala
N-CH1-CH2	30	109.0	Hyp, Pro
N-CH1-C	60	108.6	Hyp, Pro
CH1-CH2-OH1	50	109.3	Ser
CH1-CH2-NH2	50	109.6	at
CH1-CH2-CH1	30	113.2	Hyp, Leu
CH2-CH2-NH1	50	109.6	Arg
CH2-CH2-NH2	30	109.8	Lys, at
CH2-CH2-NH3	30	109.8	Lysh
CH2-CH2-CH1	30	111.8	Arg, Gln, Glu, Gluh, Lys, Lysh, Met, Pro
CH2-CH2-CH2	30	109.5	Arg, Lys, Lysh, Pro
CH3-CH2-CH1	30	108.8	Ile
C-CH2-NH1	90	110.5	Gly
C-CH2-NH2	50	110.5	at
C-CH2-NH3	50	110.5	at
C-CH2-CH1	50	110.4	Asn, Asp, Asph, His, Phe, Trp, Tyr
C-CH2-CH2	30	110.4	Gln, Glu, Gluh, Heme
S-CH2-CH1	50	117.2	Cys
S-CH2-CH2	50	110.2	Met
N-CH2-CH1	50	109.6	Hyp
N-CH2-CH2	50	109.6	Pro
SH1-CH2-CH1	50	117.2	Cysh
OH1-C-O	60	124.5	Asph, Gluh, ct
OM-C-OM	50	129.0	Asp, Glu, ct
NH1-C-O	60	124.5	all residues but Hyp, Pro
NH2-C-O	60	120.6	Asn, Gln
CH1-C-O	50	122.3	all residues but Gly
CH1-C-OH1	35	117.5	ct
CH1-C-OM	40	118.0	ct
CH1-C-NH1	35	117.5	all residues but Gly, Hyp, Pro
CH2-C-O	50	123.5	Asn, Asph, Gln, Gluh, Gly
CH2-C-OH1	40	115.9	Asph, Gluh, ct
CH2-C-OM	40	115.9	Asp, Glu, Heme, ct
CH2-C-NH1	35	115.0	Gly, Hisa, Hish
CH2-C-NH2	50	116.0	Asn, Gln
CH3-C-O	40	120.4	act
CH3-C-NH1	25	115.7	act
C-C-NH1	50	108.0	Trp
C-C-CH2	50	126.2	Trp, Heme
C-C-C	60	126.5	Trp, Heme
CR1-C-OH1	50	118.3	Tyr
CR1-C-NH1	50	126.1	Hisa, Hish, Trp
CR1-C-CH2	50	122.6	His, Phe, Trp, Tyr
CR1-C-C	50	125.6	Trp, Heme
CR1-C-CR1	50	119.0	Phe, Tyr
N-C-O	60	124.5	Hyp, Pro
N-C-CH1	35	115.7	Hyp, Pro
N-C-CH2	35	115.7	Hyp, Pro
N-C-CH3	35	115.7	act
NR-C-CH2	35	115.0	Hisb
NR-C-CR1	50	126.1	Hisb
NC2-C-NH1	60	117.0	Arg
NC2-C-NC2	60	120.3	Arg
NH1-CR1-NH1	50	108.7	Hish
C-CR1-NH1	50	109.7	Hisb, Hish, Trp
CR1-CR1-C	50	121.3	Phe, Trp, Tyr
CR1-CR1-CR1	50	123.5	Phe, Trp
NR-CR1-NH1	50	108.7	Hisa, Hisb
NR-CR1-C	50	106.9	Hisa
CH3-S-CH2	50	97.2	Met
S-S-CH2	50	104.2	Cys-Cys
CH2-N-CH1	70	117.4	Hyp, Pro
CH3-N-CH1	70	117.4	mt
CH3-N-CH2	70	117.4	mt
C-N-CH1	60	119.1	Hyp, Pro
C-N-CH2	60	119.1	Hyp, Pro
CR1-NR-C	50	108.5	Hisb
CR1-NR-CR1	50	108.5	Hisa

[a] See footnote of Table I.

constraints; eight bond angles in the four Pro residues have been omitted.

3. Computational Procedure

Three different molecular dynamics (MD) simulations have been performed: one without any constraints, denoted by the symbol NC, one with bond-length constraints, denoted by the symbol LC, and one with bond-length and bond-angle constraints, denoted by the symbol LAC. Vacuum boundary conditions have been used; that is, no external solvent molecules were included and no periodicity has been applied. The Verlet algorithm[38] has been employed to integrate the equations of motion in Cartesian coordinates. It is one of the most accurate, stable and yet simple and efficient algorithms presently available for macromolecular systems.[26] In the NC run, an integration time step $\Delta t = 10^{-3}$ "ps" has been used, as in ref 10–12. Throughout this paper we use as time unit a "ps" $(2/(4.184)^{1/2}$ ps = 0.98 ps); this is a natural unit when expressing energy in kcal mol^{-1}, mass in atomic mass units (u), and length in Å; temperature is expressed in K. These units are used throughout this paper, unless it is explicitly stated otherwise. For this time step, the root-mean-square fluctuation of the total energy (rms ΔE_{tot}) over 1 ps is

Table IV
Improper Dihedral Assignment[a]

residue	improper dihedral	residue	improper dihedral
all residues	C^n-$C_\alpha{}^n$-N^{n+1}-O^n	Phe, Tyr	C_γ-C_{δ_2}-C_{ϵ_2}-C_ζ
all residues but Gly	C_α-N-C-C_β		C_{δ_1}-C_{ϵ_1}-C_ζ-C_{ϵ_2}
Arg	C_ζ-N_{η_1}-N_{η_2}-N_ϵ		C_{δ_2}-C_{ϵ_2}-C_ζ-C_{ϵ_1}
Asn	C_γ-O_{δ_1}-N_{δ_2}-C_β	Tyr	C_ζ-C_{ϵ_1}-C_{ϵ_2}-O_η
Asp	C_γ-O_{δ_1}-O_{δ_2}-C_β	Pro	N^n-C^{n-1}-$C_\alpha{}^n$-$C_\delta{}^n$
Gln	C_δ-O_{ϵ_1}-N_{ϵ_2}-C_γ	Thr	C_β-O_{γ_1}-C_{γ_2}-C_α
Glu	C_δ-O_{ϵ_1}-O_{ϵ_2}-C_γ	Trp	C_γ-C_{δ_1}-C_{δ_2}-C_β
His	C_γ-N_{δ_1}-C_{δ_2}-C_β		C_{δ_2}-C_γ-C_{δ_1}-N_{ϵ_1}
	C_{δ_2}-C_γ-N_{δ_1}-C_{ϵ_1}		C_{δ_1}-C_γ-C_{δ_2}-C_{ϵ_2}
	N_{δ_1}-C_γ-C_{δ_2}-N_{ϵ_2}		C_γ-C_{δ_1}-N_{ϵ_1}-C_{ϵ_2}
	C_γ-N_{δ_1}-C_{ϵ_1}-N_{ϵ_2}		C_γ-C_{δ_2}-C_{ϵ_2}-N_{ϵ_1}
	C_γ-C_{δ_2}-N_{ϵ_2}-C_{ϵ_1}		C_{δ_1}-N_{ϵ_1}-C_{ϵ_2}-C_{δ_2}
	C_{δ_2}-N_{ϵ_2}-C_{ϵ_1}-N_{δ_1}		C_{δ_2}-C_{ϵ_2}-C_{ϵ_3}-C_γ
Hyp	C_γ-C_{δ_2}-C_β-O_{δ_1}		C_{ϵ_2}-C_{δ_2}-C_{ζ_2}-N_{ϵ_1}
	N^n-C^{n-1}-$C_\alpha{}^n$-$C_{\delta_2}{}^n$		C_{ϵ_3}-C_{δ_2}-C_{ϵ_2}-C_{ζ_2}
Ile	C_β-C_{γ_1}-C_{γ_2}-C_α		C_{δ_2}-C_{ϵ_2}-C_{ζ_2}-C_{η_2}
Leu	C_γ-C_{δ_2}-C_{δ_1}-C_β		C_{ϵ_2}-C_{δ_2}-C_{ϵ_3}-C_{ζ_3}
Phe, Tyr	C_γ-C_{δ_1}-C_{δ_2}-C_β		C_{ϵ_2}-C_{ζ_2}-C_{η_2}-C_{ζ_3}
	C_{δ_2}-C_γ-C_{δ_1}-C_{ϵ_1}		C_{δ_2}-C_{ϵ_3}-C_{ζ_3}-C_{η_2}
	C_{δ_1}-C_γ-C_{δ_2}-C_{ϵ_2}		C_{ϵ_3}-C_{ζ_3}-C_{η_2}-C_{ζ_2}
	C_γ-C_{δ_1}-C_{ϵ_1}-C_ζ	Val	C_β-C_{γ_2}-C_{γ_1}-C_α

[a] The atoms are named following the IUPAC-IUB convention.[39] When necessary, superscripts denote the residue number in the protein.

Table V
Dihedral Assignment[a]

residue	dihedral	residue	dihedral
all residues	C^{n-1}-N^n-$C_\alpha{}^n$-C^n	His	C_α-C_β-C_γ-N_{δ_1}
	N^n-$C_\alpha{}^n$-C^n-N^{n+1}	Hyp	C_α-C_β-C_γ-C_{δ_2}
	$C_\alpha{}^n$-C^n-N^{n+1}-$C_\alpha{}^{n+1}$		C_β-C_γ-C_{δ_2}-N
all residues but Ala, Cys, Gly, Ile, Ser, Thr, Val	N-C_α-C_β-C_γ		C_γ-C_{δ_2}-N-C_α
		Ile, Val	N-C_α-C_β-C_{γ_1}
Arg, Gln, Glu, Lys, Pro	C_α-C_β-C_γ-C_δ	Ile	C_α-C_β-C_{γ_1}-C_δ
Arg	C_β-C_γ-C_δ-N_ϵ	Leu, Phe, Trp, Tyr	C_α-C_β-C_γ-C_{δ_1}
	C_γ-C_δ-N_ϵ-C_ζ		
	C_δ-N_ϵ-C_ζ-N_{η_1}	Lys	C_β-C_γ-C_δ-C_ϵ
Asn, Asp	C_α-C_β-C_γ-O_{δ_1}		C_γ-C_δ-C_ϵ-N_ζ
Cys	N-C_α-C_β-S_γ	Met	C_α-C_β-C_γ-S_δ
Cys-Cys	$C_\alpha{}^n$-$C_\beta{}^n$-$S_\gamma{}^n$-$S_\gamma{}^m$		C_β-C_γ-S_δ-C_ϵ
	$C_\beta{}^n$-$S_\gamma{}^n$-$S_\gamma{}^m$-$C_\beta{}^m$	Pro	C_β-C_γ-C_δ-N
	$S_\gamma{}^n$-$S_\gamma{}^m$-$C_\beta{}^m$-$C_\alpha{}^m$		C_γ-C_δ-N-C_α
Gln, Glu	C_β-C_γ-C_δ-O_{ϵ_1}	Ser	N-C_α-C_β-O_γ
		Thr	N-C_α-C_β-O_{γ_1}

[a] See footnote of Table IV.

about a factor 100 smaller than that of the kinetic energy (rms ΔE_{kin}). The required computer time (NC) amounts to about 110 min of CPU time per picosecond on a VAX 11-780.

In the LC and LAC runs the constraints have been conserved by applying the SHAKE method.[25,26] In this method the relative accuracy tolerance (tol) to which the constraints are to be satisfied geometrically must be specified. However, we note that the dynamical accuracy of SHAKE depends not only on tol but also on Δt. For the LC run the time step has been taken as large as $\Delta t = 0.004$ "ps", with tol = 10^{-4}, which still yields a reasonable accuracy (rms $\Delta E_{tot} \approx$ rms $\Delta E_{kin}/30$), taking only about 37 min of CPU time per picosecond. When bond-angle constraints are included, SHAKE converges very slowly, or even diverges for the values of Δt and tol given above. Since we already had strong indications that the application of bond-angle constraints is not physically justified, only one LAC run had to be performed in order to determine the effects of rigid bond angles. It was decided, therefore, not to modify the computer codes, but to solve the convergence problem by brute force, that is, by reducing the time step Δt and relaxing the geometrical accuracy tol somewhat. In order for SHAKE to converge with the same accuracy as in the LC run, the time step would have had to be smaller than 10^{-5} ps, rendering the LAC run intolerably time-consuming. As a compromise between accuracy and efficiency we chose $\Delta t = 10^{-4}$ "ps" and tol = 3×10^{-4}. With these values the LAC runs takes about 18 h of CPU time per picosecond and yields rms $\Delta E_{tot} \approx$ rms $\Delta E_{kin}/1.5$. Evidence exists that the MD statistical averages are not much distorted by allowing such a low accuracy.[39] However, the fluctuations of various quantities will contain a component resulting from the noise generated by the lower accuracy in inte-

Table VI
Sinusoidal Dihedral Angle Parameters[a]

dihedral type	K_ϕ, kcal mol^{-1}	n	δ, deg	occurring in
X-CH1-NH1-Y	0.3	3	180	all residues but Gly, Hyp, Pro
X-CH1-CH1-Y	0.5	3	0	Ile, Thr, Val
X-CH2-NH1-Y	0.3	3	180	Arg, Gly
X-CH2-CH1-Y	0.5	3	0	all residues but Ala, Gly, Thr, Val
X-CH2-CH2-Y	0.5	3	0	Arg, Gln, Glu, Gluh, Lys, Lysh, Met, Pro, Heme
X-C-NH1-Y	7.0	2	180	all residues but Hyp, Pro
X-C-CH1-Y	0.1	3	0	all residues but Gly
X-C-CH2-Y	0.1	3	0	Asn, Asp, Asph, Gln, Glu, Gluh, Gly, His, Phe, Trp, Tyr, Heme
X-S-CH2-Y	0.5	2	0	Cys-Cys, Met
X-S-S-Y	4.0	2	0	Cys-Cys
X-N-CH1-Y	0.3	3	180	Hyp, Pro
X-N-CH2-Y	0.3	3	180	Hyp, Pro
X-N-C-Y	7.0	2	180	Hyp, Pro

[a] See footnotes of Tables I and IV.

Table VII
Hydrogen-Bond Parameters

donor atom type	acceptor atom type	E_{min}, kcal mol^{-1}	R_{min}, Å
OH1, OH2	O, OH1, OM, NP, OH2, NR	−3.5	2.80
NH1	O, NP, NR	−3.0	2.95
	OH1, OH2	−3.0	3.08
	OM	−3.0	3.10
NH2, NC2	O, OH1, OM, NP, OH2, NR	−2.5	2.87
NH3	O, OH1, OM, NP, OH2, NR	−2.5	3.00

Table VIII
Bond Angles Excluded from Bond-Angle Constraints[a]

residue	bond angle	residue	bond angle
His	C_{δ_2}-N_{ϵ_2}-C_{ϵ_1}	Pro	C_β-C_γ-C_δ
Hyp	C_β-C_γ-C_{δ_2}		C^{n-1}-N^n-$C_\delta{}^n$
		Trp	C_{δ_1}-C_γ-C_{δ_2}
	C^{n-1}-N^n-$C_{\delta_2}{}^n$		N_{ϵ_1}-C_{ϵ_2}-C_{ζ_2}

[a] See footnote of Table IV.

grating the equations of motion. Also, no metric tensor potential[15] has been included in the energy function, which is justified for the LC run but may not be for the LAC run.[40,41] This should be kept in mind when interpreting the LAC results.

The computer time per step for the runs is determined mainly by the time required for scanning all of the atom pairs in the protein when calculating the nonbonded interaction. When making the pair list, we use a cutoff radius R_l = 8.5 Å, which is somewhat larger than the cutoff radius for the nonbonded interaction, R_c = 8.0 Å. This enables us to update the pair list only every 0.02 "ps" (NC: 20 steps; LC: 5 steps; LAC: 200 steps), since an atom will certainly not move more than 0.5 Å in 0.02 "ps". The hydrogen-bond pair list is kept fixed during the runs.

Before a system may be considered to be in equilibrium, it needs time to "age". The equilibration time depends mainly on the range of the interactions, the system size and the initial configuration, and velocities. Various equilibration procedures have been used in MD simulations. For our system, we started with the X-ray structure[37] and minimized the total potential energy; i.e., by use of the method of steepest descents, the energy was minimized until the decrease was less than 10^{-3} kcal mol^{-1} per step, which occurred after 157 steps. At that point the step size was as small as 0.001 Å and the root-mean-square gradient (mean over all atoms) was 1.3 kcal mol^{-1} Å^{-1}. This is done to eliminate the strain present in the X-ray configuration. The initial atomic velocities are taken from a Maxwellian distribution with a temperature (450 K) about one and a half times as high as the reference temperature T_0 (300 K) at which the system is required to settle down, and the MD run is started. We note, however, that it is probably better to start the run in a smoother way, by raising the temperature from 0 K to the desired value in a series of small intervals followed by some MD equilibration steps. The conversion of part of the initial kinetic energy into potential energy is a fast process, requiring less than tenths of picoseconds. For example, the average temperature was as low as 247 K over the first 0.1 "ps" of the run. Since the starting configuration of the protein is generally not near a global minimum of the potential energy, the temperature will rise when the run proceeds, as the protein reaches regions of lower potential energy in configuration space. However, the temperature rise decreases when the system ages. This process is monitored by calculating the averages of the temperature and of the various contributions to the potential energy over 0.1-"ps" intervals during the first "picosecond" and over 1-"ps" intervals for the remaining part of the run. In order to keep the protein near the reference temperature, T_0, during the equilibration, all velocities are rescaled by a factor

$$[2T_0/\langle T\rangle - 1]^{1/2} \tag{13}$$

whenever

$$|\langle T\rangle - T_0| > \Delta T_0, \qquad \Delta T_0 = 10 \text{ K} \tag{14}$$

We initially chose ΔT_0 = 5 K, but this value triggered too many rescalings since the root-mean-square fluctuation of the temperature turned out to be about 8 K (NC). After 25 "ps" no rescaling was required. The NC run covered 55 "ps", of which the final 25 "ps" has been used for analysis, the first 30 "ps" being the equilibration period. This period is long for proteins, because of the high connectivity of the covalently bonded system and the long range of the electrostatic potential. From Table IX it is observed that the 1-"ps" averages of the temperature and the various potential energy terms form oscillating, but reasonably stable sequences. The protein contracts during the initial stages of the run because of the vacuum boundary condition, as can be concluded from the increasing number of nonbonded pairs (R_c = 8.5 Å) in the molecule.

The bond-length constraints were taken equal to the average bond lengths, as obtained from the analysis period of the NC run. A long equilibration period has been avoided for the LC run by starting from the NC configuration and velocities at 20 "ps"; that is, for this configuration the bond lengths were frozen at the required values and the velocity components along the directions of the constraints were removed. The LC run covered 40 "ps", of which the final 25 "ps" has been used for analysis. During the 15-"ps" equilibration period the velocities were rescaled only once (ΔT_0 = 10 K). The stability of the various energy contributions can be judged from Table IX.

In the case of bond-length and bond-angle constraints, SHAKE (tol = 10^{-3}) did not converge completely within 300 interactions when trying to fit the NC configuration at 20 "ps" to the average bond lengths and bond angles of the NC run. Since the differences were very small, we took the bond lengths and bond angles of the shaken configu-

Table IX
Average Temperature and Energies as a Function of Time [a]

method	period, "ps"	$\langle T \rangle$	$\langle E_{kin} \rangle$	$\langle E_{pot} \rangle$	$\langle E_b \rangle$	$\langle E_\theta \rangle$	$\langle E_\xi \rangle$	$\langle E_\phi \rangle$	$\langle E_{hb} \rangle$	$\langle E_{el} \rangle$	$\langle E_{vdw} \rangle$	nonbonded pairs
X-ray				−58.2	69.3	110.8	6.4	67.7	−119.8	−34.7	−157.9	18 287
EM				−360.4	21.6	81.3	5.1	62.3	−133.2	−82.4	−315.0	18 382
NC	4-5	307	417.5	−236.5	146.3	211.0	43.0	92.3	−114.8	−257.6	−356.8	20 372
	9-10	304	413.9	−282.0	152.4	216.0	44.0	97.5	−118.0	−285.5	−388.3	20 842
	14-15	302	411.1	−284.0	156.0	210.3	45.3	95.2	−120.7	−281.4	−388.6	20 801
	19-20	301	409.4	−297.8	155.0	205.4	44.9	96.8	−114.5	−294.2	−391.2	21 130
	24-25	303	412.4	−300.8	159.9	206.3	43.6	95.0	−117.6	−296.0	−391.9	21 082
	29-30	299	405.6	−308.2	163.9	201.7	41.9	99.5	−116.1	−299.2	−399.8	21 226
	34-35	297	404.0	−306.6	163.3	205.2	43.4	95.9	−119.9	−291.6	−403.3	20 999
	39-40	300	407.6	−310.2	163.3	206.9	44.3	93.5	−120.4	−295.1	−402.7	21 158
	44-45	300	408.2	−310.9	166.3	204.6	43.3	94.6	−120.4	−295.0	−404.2	21 117
	49-50	302	410.7	−313.3	167.7	207.2	42.5	96.4	−119.6	−304.5	−403.1	21 260
	54-55	303	411.4	−314.0	161.3	208.5	47.6	95.8	−120.5	−304.1	−402.6	21 213
LC	4-5	299	267.1	−463.5		211.6	45.7	94.2	−118.8	−307.4	−388.9	20 871
	9-10	296	265.1	−464.5		209.1	43.4	97.3	−115.8	−303.6	−394.9	21 305
	14-15	303	271.0	−473.2		216.4	46.7	96.5	−115.7	−314.7	−402.3	21 586
	19-20	308	275.6	−481.0		219.0	48.2	101.9	−117.2	−325.9	−407.1	21 399
	24-25	306	273.4	−482.0		215.6	45.3	100.8	−115.2	−317.4	−411.1	21 266
	29-30	307	274.6	−485.9		213.6	43.9	98.8	−117.3	−318.1	−406.7	21 622
	34-35	306	273.7	−488.1		210.4	44.6	99.3	−117.7	−319.0	−405.7	21 208
	39-40	301	269.6	−486.9		210.7	43.2	102.5	−117.7	−321.1	−404.4	21 354
LAC	4-5	268	71.9	−758.0				89.2	−124.2	−295.1	−427.9	21 299
	9-10	288	77.2	−770.0				92.4	−123.8	−309.3	−429.2	21 518
	12-13	312	83.8	−785.3				88.5	−126.8	−308.6	−438.3	21 385
	17-18	294	78.8	−793.2				93.2	−126.8	−318.4	−441.0	21 568
	22-23	299	80.2	−791.6				92.5	−126.0	−318.6	−439.5	21 766
	27-28	306	82.2	−789.8				88.4	−126.2	−310.6	−441.4	21 537

[a] The values after energy minimization are denoted by the symbol EM. The time intervals over which the quantities are averaged are in "ps", the temperature (T) in K, and the energies in kcal mol^{-1}. These are the kinetic energy (E_{kin}), potential energy (E_{pot}), covalent bond energy (E_b), bond-angle energy (E_θ), improper dihedral energy (E_ξ), dihedral energy (E_ϕ), hydrogen-bond energy (E_{hb}), electric energy (E_{el}), and the van der Waals energy (E_{vdW}). The last column contains the number of nonbonded pairs within 8.5 Å for the final configuration of the averaging period or for a previous one.

ration as contraints for the LAC run. This shaken configuration was taken as the starting point for the LAC run and, as above, the velocity components along the constraint directions were eliminated. When SHAKE is applied with values for Δt and tol that are too large, as is the case for the LAC run, it causes the total energy to drop; that is, energy leaks out of the system. Therefore, the velocities are rescaled upward with a factor of about 1.02 approximately every 0.5 "ps" during the whole 28 "ps" of the LAC run. The final 15 "ps" has been used for analysis. The variability of the different energy terms during this period can be seen in Table IX.

The requirement for scaling of the velocities in the NC and LC runs has an origin different from that in the LAC run. Even if the total energy is perfectly conserved (NC, LC), the protein may "fall" into a region of low potential energy, causing the temperature to rise. When this happens, rescaling is performed to keep the temperature close to the reference value, T_0, but only during the equilibration period of the runs. In the LAC run, the potential energy is approximately constant, but energy is leaking out of the system due to the inaccuracy of SHAKE and this causes the temperature to decrease. Thus, rescaling is required during the whole run to keep the temperature near the reference value.

Taking the initial atom velocities from a Maxwellian distribution (NC run) generally gives the molecule an overall translational motion and an overall rotational motion about its center of mass. Because of the vacuum boundary condition, these motions are conserved during the run. It is convenient to stop them at the start of the run because this simplifies the comparison of different configurations. The finite accuracy of the integration algorithm and SHAKE will produce a slight overall translational and rotational motion of the molecule in the course of a run. Therefore, these motions are routinely halted every 5 "ps" or when the velocities are rescaled, even if they are negligibly small.

All the averages and correlation functions reported in the next sections are calculated from the last 25-"ps" (NC, LC) or 15-"ps" (LAC) portions of the runs; coordinates, velocities, and energies were stored every 0.01 "ps" (NC, LAC) or 0.008 "ps" (LC). Besides averages over the total analysis period of 25 "ps" (LAC: 15 "ps"), averages over 5 "ps" have also been obtained; the latter bring out the dependence of the results on the time span over which the system is being averaged and remove the effect of a slow drift in the overall structure from the average value.

4. Molecular Dynamics without Constraints (NC)

In Table X the average values of the temperature, $\langle T \rangle$, and its root-mean-square fluctuation, $\langle (\Delta T)^2 \rangle^{1/2}$, are given. Throughout this paper a time average is denoted by angular brackets and the symbol Δ is used to denote the difference between the actual value and the average value of a quantity; that is, $\Delta T \equiv T - \langle T \rangle$. An overbar denotes a sum over Cartesian components (if possible) and subsequent averaging over atoms with a given name. The temperature is obtained from the total kinetic energy E_{kin} by assuming $^1/_2 kT$ kinetic energy per degree of freedom, where k is Boltzmann's constant. The classical heat capacity per atom, c_{at}, is calculated from[42]

$$c_{at} = \{2/(\text{ndf}) - \langle (\Delta E_{kin})^2 \rangle / \langle E_{kin} \rangle^2\}^{-1} \cdot N_{at}^{-1} \cdot k \quad (15)$$

The number of degrees of freedom (ndf) of the system is given in Table X, and the number of atoms $N_{at} = 458$. Not

Table X
Temperature and Heat Capacity [a]

period, "ps"	NC (ndf = 1368) ⟨T⟩, K	NC ⟨(ΔT)²⟩^1/2, K	NC c_{at}, k	LC (ndf = 900) ⟨T⟩, K	LC ⟨(ΔT)²⟩^1/2, K	LC c_{at}, k	LAC (ndf = 278) ⟨T⟩, K	LAC ⟨(ΔT)²⟩^1/2, K	LAC c_{at}, k
0-5	295.5	8.4	3.37	305.5	10.8	2.23 (3.38)	291.8	12.0	0.40 (1.95)
5-10	299.1	8.5	3.34	306.2	10.6	2.14 (3.26)	293.3	12.4	0.40 (1.99)
10-15	300.9	8.0	2.90	304.0	10.2	2.01 (3.05)	296.5	11.7	0.39 (1.91)
15-20	302.2	8.6	3.35	305.6	10.5	2.09 (3.18)			
20-25	300.7	8.0	2.91	304.4	10.7	2.21 (3.37)			
0-25 (15)	299.7	8.6	3.45	305.2	10.6	2.15 (3.27)	293.9	12.2	0.40 (1.96)
ref 11	295	6.9	2.3						
ref 12	306	8.5	3.14						

[a] The heat capacity given between parentheses is obtained by multiplication by the ratio of the number of degrees freedom without constraints and with constraints.

Table XI
Average Kinetic Energy per Atom Name [a]

atom name	no. of atoms	NC ⟨E_{kin}⟩, kcal mol^{-1}	NC ndf	NC $\overline{\langle T \rangle}$, K	LC ⟨E_{kin}⟩, kcal mol^{-1}	LC ndf	LC $\overline{\langle T \rangle}$, K	LAC ⟨E_{kin}⟩, kcal mol^{-1}	LAC ndf	LAC $\overline{\langle T \rangle}$, K
all	458	0.889	3	300	0.596	906/458	305	0.177	284/458	293
N	58	0.889	3	300	0.601	[1 × 5/2 + 4 × 3/2 + 53 × 2]/58	308	0.145		
C_α	58	0.904	3	305	0.439	[6 × 2 + 52 × 3/2]/58	287	0.121		
C_{ϵ_1}	8	0.917	3	309	0.568	2	288	0.182		
N_{ζ_1}	6	0.862	3	290	0.746	5/2	302	0.313		

[a] The temperature is calculated from the formula ⟨T⟩ = ⟨E_{kin}⟩/[(ndf)/2k] f, where the factor f accounts for the fact that the molecule has no translational or rotational motion; that is, f = 1374/1368 (NC), 906/900 (LC), or 284/278 (LAC). The symbol ndf denotes the number of degrees of freedom per atom. In the LC case each bond-length constraint is assumed to reduce the number of degrees of freedom of each of the two connected atoms by $^1/_2$. So, the N atoms generally have ndf = 2, except in the four Pro residues, where ndf = $^3/_2$, and at the amino terminal of the chain, where ndf = $^5/_2$. The C_α atoms have ndf = $^3/_2$, except for the six Gly residues, where ndf = 2.

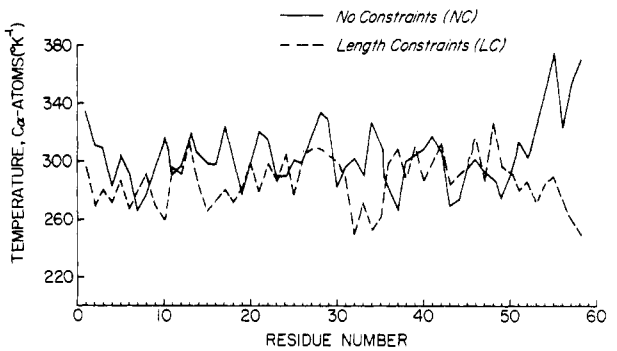

Figure 1. Average temperature of C_α atoms over 25 "ps": (—) NC run; (---) LC run.

withstanding the 30-"ps" equilibration period, the temperature shows a slow upward drift, reflecting the fact that the protein is still moving toward regions of lower potential energy in configurational space. This explains that the calculated heat capacity is 10% larger than that of ref 11, which was obtained after an equilibration period of 75 "ps"; the too small c_{at} of ref 10 is due to averaging over too short a period (2 "ps").

In Figure 1, ⟨T⟩ for the C_α atoms is plotted as a function of residue number. The kinetic energy is equally distributed over the backbone except that it is slightly higher at the carboxy terminus. The distribution among the individual atoms changes rapidly with time, as has been checked by comparing 5-"ps" averages; they look very different. The kinetic energy of the side-chain atoms is generally not larger than that of the backbone atoms, as is illustrated in Table XI. The velocity autocorrelation functions, ⟨$\vec{v}(0)\cdot\vec{v}(t)$⟩, for the C_α atoms are rather similar. Their spectral densities are dominated by peaks at ν_1 =

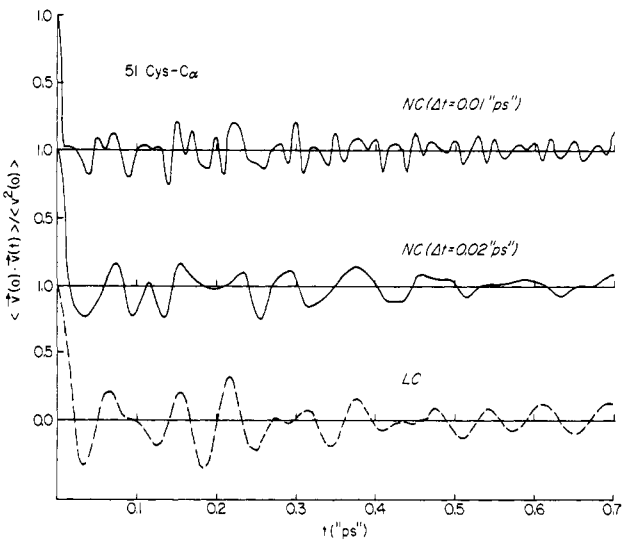

Figure 2. Velocity autocorrelation function for 51 Cys C_α. Top curve is for NC run with time intervals of 0.01 ps; middle curve is for NC run with time interval of 0.02 ps: bottom curve is for LC run with time intervals of 0.01 ps.

13 and 17 "ps"$^{-1}$. In Figure 2 the normalized velocity autocorrelation function for 51 Cys C_α is displayed as an example. The second curve is drawn with half the resolution of the first one in order to eliminate the high-frequency oscillations. The C_α atoms in more flexible regions of the protein (15 Lys, 26 Lys) show more pronounced extrema. An example of a velocity autocorrelation function for a side-chain atom (15 Lys N_ζ) can be found in ref 43.

When two protein configurations are compared, their centers of mass are made identical and subsequently one

Table XII
Fluctuations in the Atom Cartesian Coordinates[a]

atom name	period, "ps"	$\overline{\langle(\Delta x)^2\rangle}^{1/2}$, Å			$\overline{\lvert\langle(\Delta x)^3\rangle\rvert}^{1/3}$, Å			$\overline{\langle(\Delta x)^4\rangle}^{1/4}$, Å		
		NC	LC	LAC	NC	LC	LAC	NC	LC	LAC
all	0-5	0.58	0.55	0.28	0.45	0.40	0.23	0.76	0.75	0.41
	5-10	0.50	0.56	0.34	0.38	0.38	0.26	0.66	0.76	0.47
	10-15	0.52	0.57	0.31	0.40	0.49	0.23	0.74	0.88	0.44
	15-20	0.56	0.56		0.40	0.41		0.74	0.77	
	20-25	0.53	0.56		0.39	0.44		0.68	0.79	
	rms	0.54	0.56	0.31	0.40	0.43	0.24	0.72	0.80	0.44
	0-25 (15)	0.60	0.72	0.34	0.44	0.50	0.24	0.79	0.97	0.48
	X-ray	0.74								
	ref 11	0.90								
	ref 12	0.75								
C_α	0-5	0.43	0.43	0.19	0.26	0.27	0.12	0.47	0.47	0.21
	5-10	0.35	0.41	0.26	0.22	0.25	0.16	0.38	0.46	0.29
	10-15	0.37	0.40	0.24	0.25	0.29	0.15	0.42	0.45	0.27
	15-20	0.43	0.41		0.28	0.27		0.48	0.45	
	20-25	0.42	0.39		0.27	0.25		0.48	0.43	
	rms	0.40	0.41	0.23	0.26	0.27	0.14	0.45	0.45	0.26
	0-25 (15)	0.44	0.54	0.25	0.26	0.31	0.15	0.48	0.60	0.28
	X-ray	0.68								
	ref 11	0.74								
	ref 12	0.60								

[a] The symbol rms denotes the root-mean-square (cube root mean for third power; fourth root mean for fourth power) of the 5 "ps" values.

of them is rotated around its center of mass such that the mean-square difference of the C_α atom coordinates of the two configurations is at a minimum;[44] including all atoms instead only of the C_α atoms makes little difference. It is not necessary to perform this rotational least-squares fit for every configuration when time averaging, since the protein has no overall translational or rotational motion in a vacuum run. The root-mean-square deviation of the time-averaged structure from the X-ray structure amounts to 2.7 Å for all atoms and 2.1 Å for the C_α atoms. The largest deviations come from the two ends of the protein, the exposed side chains, and the extended loop (25 Ala, 26 Lys, 27 Ala, 28 Gly), all of which have a tendency to fold back onto the protein because it is in vacuo.

The atom positional fluctuations are listed in Table XII. Even after the rather long equilibration period, the fluctuations depend slightly on the time span over which they are averaged; however, the largest 5-"ps" values are close to the 25-"ps" values for the NC run. For the LC run (see below) the difference between the 5-"ps" and the 25-"ps" values is larger, even for the C_α atoms, which indicates that in this run the protein has a slow motion in which many atoms are involved on a time scale of many picoseconds. The slow variation in the number density (Table XX) also suggests that a slow global mode occurs in the LC run. The large atom positional root-mean-square fluctuations of ref 11 probably reflect the fact that the protein is drifting away from the initial X-ray structure. The fact that the 100-"ps" values of ref 12 are larger than the present values has another origin, since the run of ref 12 has been equilibrated over a period of 75 "ps". That the root-mean-square fluctuations are smaller in the present run may be due to the higher average atom density (Table XX); the latter results from the larger cutoff radius used for the nonbonded interactions (8.0 Å) than the 6.5 Å used in ref 12 and 45. The root-mean-square fluctuations of the C_α atoms are plotted in Figures 3 and 4, averaged over 25 and 5 "ps", respectively. The NC curves are very close to each other and agree qualitatively with that of ref 10. No correlation is observed between the average temperature (Figure 1)

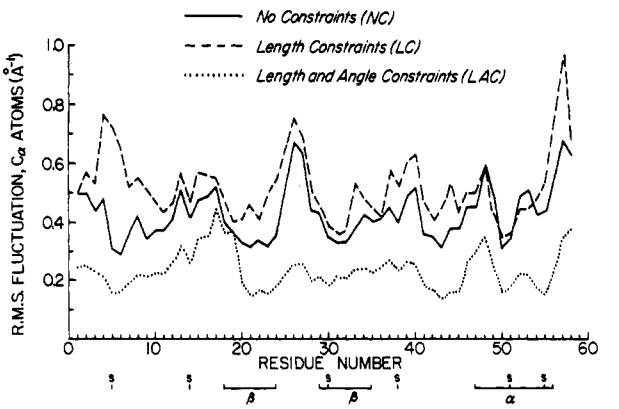

Figure 3. Root-mean-square fluctuations of C_α atoms over 25 "ps": (—) NC run; (- - -) LC run; (···) LAC run.

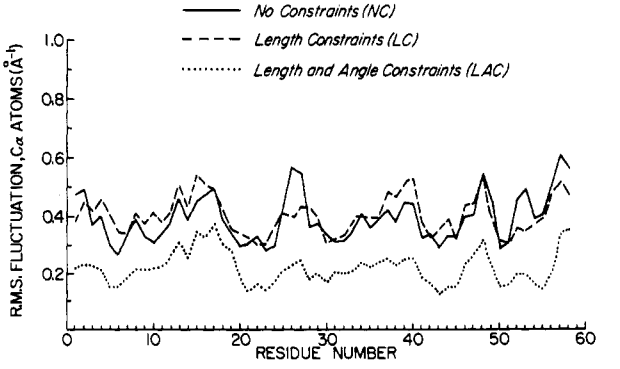

Figure 4. Root-mean-square fluctuations of C_α atoms over 25 "ps" averaged over subintervals of 5 "ps": (—) NC run; (- - -) LC run; (···) LAC run.

and the root-mean-square positional fluctuations (Figures 3 and 4) for the C_α atoms as a function of residue number. The latter curves reflect the flexibility of the various segments of the protein and could be directly compared to experimental temperature factors.[43,46,47] A positive correlation is observed between the size of C_α root-mean-

Table XIII
Root-Mean-Square Fluctuations and Anisotropies per Atom Name[a]

atom name	$\langle(\Delta x)^2\rangle^{1/2}$, Å			max ev $(\langle\Delta x_\alpha \Delta x_\beta\rangle)^{1/2}$ / min ev $(\langle\Delta x_\alpha \Delta x_\beta\rangle)^{1/2}$		
	NC	LC	LAC	NC	LC	LAC
all	0.60	0.72	0.34	2.23 (2.62)	2.51 (2.61)	2.53 (2.75)
N	0.43	0.52	0.24	1.80 (2.19)	2.30 (2.23)	1.98 (2.28)
C_α	0.44	0.54	0.25	1.83 (2.24)	2.34 (2.25)	2.01 (2.28)
C_{ϵ_1}	0.68	0.82	0.34	2.78 (2.94)	2.67 (2.84)	2.58 (2.71)
N_ζ	0.82	0.88	0.47	2.13 (2.68)	2.92 (2.78)	2.57 (2.56)

[a] Time averages over 25 "ps". The mean of the ratios obtained by averaging over 5 "ps" is given between parentheses.

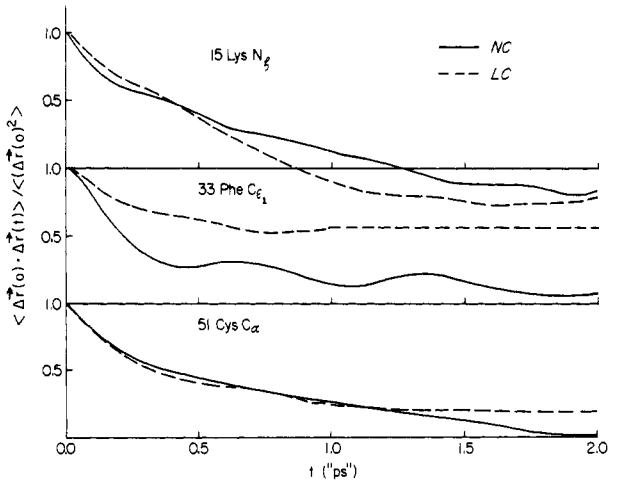

Figure 5. Position autocorrelation function for sample atoms: (—) NC run; (---) LC run.

square fluctuations and the root-mean-square deviations of the time-averaged C_α positions from the X-ray structure.[37] The mobility of the side-chain atoms is larger than that of the backbone atoms (Table XIII). The largest fluctuations (1.2–1.8 Å) are found in side chains sticking out from the protein surface; 29 Leu, 39 Arg, 49 Glu, 53 Arg, 52 Met, 17 Arg, and 26 Lys. A more detailed analysis of the atom positional fluctuations and the deviations from the X-ray structure is given in ref 43. The anisotropy of the positional fluctuations is measured by computing the ratio of the largest (max) and the smallest (min) eigenvector (ev) of the time-averaged fluctuation tensor with components $\langle\Delta x_\alpha \Delta x_\beta\rangle$ ($\alpha, \beta = 1, 2, 3$). They are sensitive to the averaging time span; the 25-"ps" values differ from the means of the 5-"ps" values, given in parentheses in Table XIII. The values should be considered only as rough estimates of the atom anisotropies. On the average, the largest components of the atom fluctuations are 2 or 3 times as large as the smallest, as in ref 11. The positional distribution functions deviate significantly from a Gaussian form, as can be concluded from the nonzero values for the skewness in Table XII.

Selected positional fluctuation autocorrelation functions $C(t) \equiv \langle\Delta\vec{r}(t)\cdot\Delta\vec{r}(0)\rangle$ are displayed in Figure 5. They generally decay within a few "picoseconds", unless the atoms are part of a segment of the protein that exhibits a slow motion; e.g., the aromatic ring of 33 Phe shows a slow motion on a time scale of several "picoseconds" (Figure 5). A crude estimate of the short-time fluctuation relaxation times is obtained by assuming that

$$C(t) = (1 - A)e^{-t/\tau_e} + A \qquad (16)$$

The values of τ_e given in Table XIV are computed from

$$C(0.1) = [1 - C(2.0)]e^{-0.1/\tau_e} + C(2.0) \qquad (17)$$

The longer relaxation time of the atoms in 26 Lys is likely

Table XIV
Atom Fluctuation Relaxation Times

residue	atom	τ_e, "ps" NC	τ_e, "ps" LC	atom	τ_e, "ps" NC	τ_e, "ps" LC
15 Lys	C_α	0.70	0.96	N_ζ	0.37	0.51
26 Lys	C_α	1.46	0.57	N_ζ	1.24	0.43
33 Phe	C_α	0.55	0.59	C_{ϵ_1}	0.42	0.35
36 Gly	C_α	0.34	0.27			
43 Asn	C_α	0.33	0.45	O_{δ_1}	0.13	0.19
51 Cys	C_α	0.42	0.32	S_γ	0.41	0.52

to be due to the fact that the whole extended loop (residues 25–28) can move relatively freely. The 43 Asn O_{δ_1} atom is hydrogen bonded to 23 Tyr N, which is part of the relatively rigid β sheet; this may account for its relatively short relaxation time.

The 25-"ps" averages and fluctuations for the internal coordinates of the protein, such as bond angles, improper dihedral angles, and dihedral angles, are given in Tables XV and XVI. The time-averaged bond angles and improper dihedral angles are very close to the X-ray values; the overall root-mean-square deviation is only a few degrees. The inclusion of the improper dihedral harmonic term in the potential makes the root-mean-square fluctuations of the bond angles smaller than in ref 10, where no improper torsions were applied. A bond-angle autocorrelation function is displayed in Figure 6 together with its spectral density. Only the high-frequency range is shown, since the statistical accuracy decreases with decreasing frequency. The correlation function shows irregular oscillatory behavior which decays strongly during the first 0.1 "ps". This suggests that the bond-angle motion is rather strongly coupled to the other degrees of freedom of the molecule, so that the application of bond-angle constraints will significantly alter the dynamics of the protein. The application of bond-length constraints should be valid since the observed bond-length fluctuations show the pattern expected for a system of strongly coupled local oscillators that is weakly coupled to the rest of the protein, as has been pointed out in ref 10. The autocorrelation function for the peptide dihedral angle ω decays more slowly than that for the bond angles, but its main frequency is not far from that of the bond-angle spectral density (Figure 6). The fluctuation in the dihedral angles amounts to 20° for all the dihedral angles and to about 15° for the ϕ and ψ angles. The latter are shown in Figure 7 as a function of residue number. No correlation is observed between the C_α atom positional root-mean-square fluctuations (Figure 3) and the ϕ,ψ root-mean-square fluctuations (Figure 7). The fluctuations of ϕ_n are strongly coupled to those of ψ_{n-1}, where n denotes the residue number (see also Table XVII). The large peaks in the curves represent dihedral angles that seem to possess more than one favorable conformation. For example, the 45 Phe ψ,46 Lys ϕ pair shows the largest fluctuations. Their 5-"ps" aver-

Table XV
Bond-Angle and Improper Dihedral Averages and Fluctuations[a]

	$\overline{\langle q \rangle}$		$\overline{q(X)}$	$\overline{(q_1 - q_2)^2}^{1/2}$			$\overline{\langle(\Delta q)^2\rangle}^{1/2}$		$\overline{\langle(\Delta q)^3\rangle}^{1/3}$		$\overline{\langle(\Delta q)^4\rangle}^{1/4}$	
	NC	LC	X-ray	NC/ X-ray	LC/ X-ray	NC/ LC	NC	LC	NC	LC	NC	LC
bond angles												
C-N-C_α	122.3	122.3	121.0	2.0	2.0	0.8	3.4	3.5	1.4	1.4	4.5	4.6
N-C_α-C	111.4	111.3	111.5	5.5	5.8	1.3	3.4	3.5	1.1	1.1	4.5	4.6
C_α-C-N	114.7	114.4	115.5	2.6	2.6	1.3	3.6	3.7	−1.1	−1.0	4.7	4.9
C_α-C-O	120.5	120.7	120.8	1.7	1.6	1.0	3.4	3.4	0.9	0.4	4.5	4.5
all				2.9	3.0	1.2	3.7	3.8	−0.3	−0.6	5.1	5.2
improper dihedrals												
C_α-N-C-C_β	34.0	33.9	34.9	2.6	2.6	1.3	3.4	3.5	−1.5	−1.5	4.6	4.6
C-C_α-N-O	0.0	0.0	0.0	2.9	3.0	1.5	5.6	5.8	−1.0	−1.1	7.3	7.5
all				3.0	2.9	1.4	5.0	5.1	0.6	−0.8	6.8	7.0

[a] The various internal coordinates are denoted by the symbol q. The ones calculated from the X-ray structure are denoted by $q(X)$. When two structures 1 and 2 are compared, they are listed as 1/2. Time averages are over 25 "ps". All values in degrees.

Table XVI
Dihedral Averages and Fluctuations[a]

	$\overline{\langle q \rangle}$			$\overline{q(X)}$	$\overline{(\langle q \rangle - q(X))^2}^{1/2}$			$\overline{(\langle q \rangle_1 - \langle q \rangle_2)^2}^{1/2}$		
dihedrals	NC	LC	LAC	X-ray	NC/ X-ray	LC/ X-ray	LAC/ X-ray	NC/LC	NC/LAC	LC/LAC
C-N-C_α-C					30.5	36.0	33.5	16.9	7.9	18.3
N-C_α-C-N					36.4	39.4	37.9	23.6	8.8	23.3
C_α-C-N-C_α	178.5	177.8	177.7	179.5	9.7	10.0	10.5	4.7	3.7	6.2
C_β-S_γ-S_γ-C_β					10.4	5.0	17.2	7.4	7.1	14.1
all					44.9	50.5	45.9	30.1	23.3	31.0

	$\overline{\langle(\Delta q)^2\rangle}^{1/2}$			$\overline{\langle(\Delta q)^3\rangle}^{1/3}$			$\overline{\langle(\Delta q)^4\rangle}^{1/4}$		
dihedrals	NC	LC	LAC	NC	LC	LAC	NC	LC	LAC
C-N-C_α-C	15.3	15.7	8.1	8.5	8.6	1.8	22.1	22.4	11.7
N-C_α-C-N	14.3	14.5	7.5	−7.3	−6.9	−4.2	20.7	20.3	11.1
C_α-C-N-C_α	7.7	7.9	4.5	−1.0	−2.2	−1.2	10.3	10.7	6.2
C_β-S_γ-S_γ-C_β	9.0	10.6	4.6	−3.5	−4.0	1.3	12.0	14.0	6.0
all	19.5	21.3	9.0	−12.8	15.6	3.4	38.7	44.9	18.9

[a] See footnote of Table XV.

ages, given in Table XVIII, display a 50° jump after about 5 "ps". Another example is the 23 Tyr ψ,24 Asn ϕ pair in the LC run, which shows a 40° jump after about 15 "ps". The 17 Arg ψ,18 Ile ϕ pair shows two jumps of 13° and 25° after 10 and 20 "ps", respectively. For this pair, the ϕ,ψ cross-correlation function is shown in Figure 8. It decays exactly as the autocorrelation function of the ϕ angle, as expected from the large ϕ,ψ correlation coefficient of −0.91 (Table XVII). Apart from transitions between neighboring conformational minima, the dihedral correlation functions are highly damped. In Table XVII various correlation coefficients are given. No strong correlation between neighboring bond angles was found. In the backbone, only the (ψ_{n-1},ϕ_n) pairs are strongly coupled, as observed earlier. The side-chain couplings depend on residue and environment. The χ^2 and χ^3 angles in the Pro residues are almost 100% coupled, as expected. For the Phe and Tyr residues the correlation between the two side-chain dihedral angles is stronger when the aromatic ring is buried in the protein; the exposure to solvent of the listed resides is 35% (10 Tyr), 10% (22 Phe) and 6% (35 Tyr).[43] The positive correlation coefficients in the 39 Arg side chain suggest that angular momentum is transferred between neighboring dihedral angles. Dihedral angles showing more than four transitions between the minima of the local sinusoidal potential are listed in Table XIX. Here, a dihedral

Table XVII
Bond-Angle and Dihedral Cross-Correlation Coefficients[a]

q_1	q_2	$\langle \Delta q_1 \Delta q_2 \rangle / \langle(\Delta q_1)^2\rangle^{1/2}\langle(\Delta q_2)^2\rangle^{1/2}$ NC	LC
22 Phe N-C_α-C	22 Phe C-N-C_α	0.06	0.15
	C_α-C-N	0.16	0.17
	23 Tyr C-N-C_α	−0.07	−0.08
7 Glu ϕ	6 Leu ϕ	0.07	0.08
	ψ	−0.88	−0.86
18 Ile ϕ	17 Arg ψ	−0.91	−0.79
	18 Ile ψ	−0.01	−0.05
8 Pro χ^2	χ^3	−0.96	−0.95
10 Tyr χ^1	χ^2	−0.07	−0.21
22 Phe χ^1	χ^2	−0.37	−0.35
35 Tyr χ^1	χ^2	−0.63	−0.46
39 Arg χ^1	χ^2	0.14	0.11
χ^2	χ^3	0.20	0.21
χ^3	χ^4	0.33	0.41
χ^4	χ^5	−0.01	−0.07

[a] See footnote of Table XV.

transition is defined as follows: After going through the minimum of a well of the dihedral angle potential, a transition is considered to have occurred when the value of the angle crosses the minimum of an adjacent well of

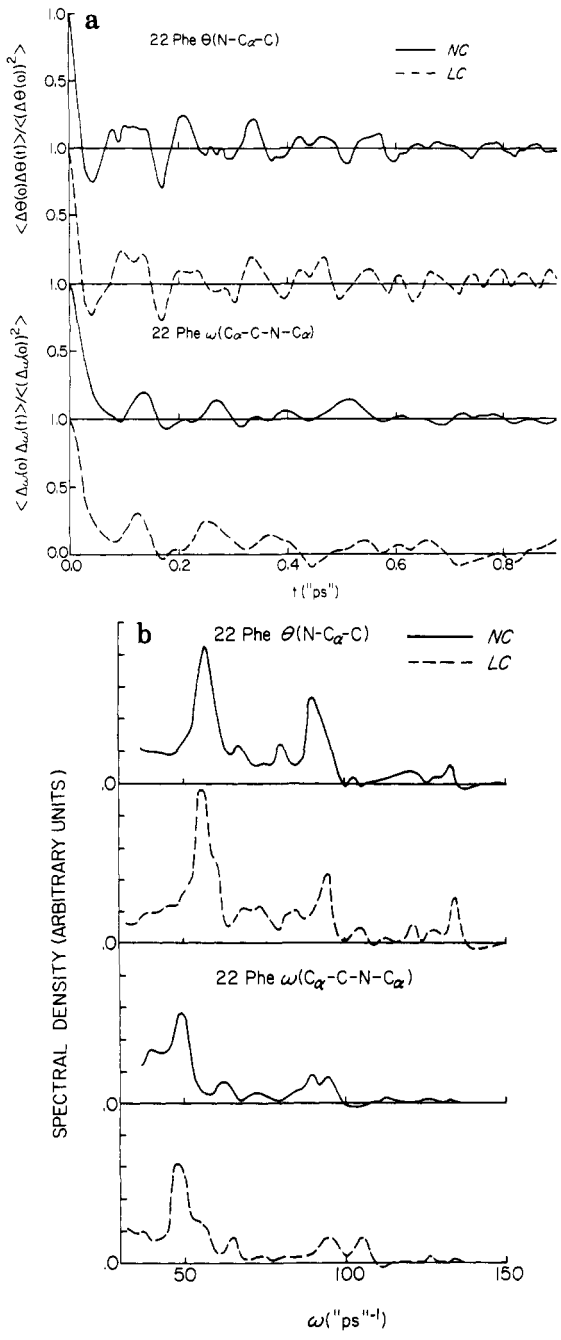

Figure 6. Bond-angle and dihedral angle autocorrelation and spectral density functions: (—) NC run; (---) LC run. (a) Autocorrelation function; (b) spectral density.

the dihedral term in the potential. With this strict definition, 110 transitions were observed in 25 "ps", of which 3 occurred in the backbone. They are one transition (0° → 120°) of 3 Asp ϕ and two transitions (−120° → 0° → 120°) of 46 Lys ϕ with an interval 3.1 "ps". The time span between two consecutive transitions in different angles ranges upward from 0.2 "ps". Most transitions occur in the side chains of residues at the surface of the protein, such as Lys and Arg. The number of transitions is too small to permit a statistical analysis of the dynamical correlation between consecutive transitions of one dihedral angle or of two neighboring dihedral angles.

To look for global protein motions of the "breathing mode" type, it is possible to follow the variation in the radius of gyration

$$R_{gyr} \equiv \{N_{at}^{-1}\sum_{i=1}^{N_{at}}[\vec{r}_i - \vec{R}_{cm}]^2\}^{1/2}; \qquad |\vec{r}_i - \vec{R}_{cm}| < r_c \quad (18)$$

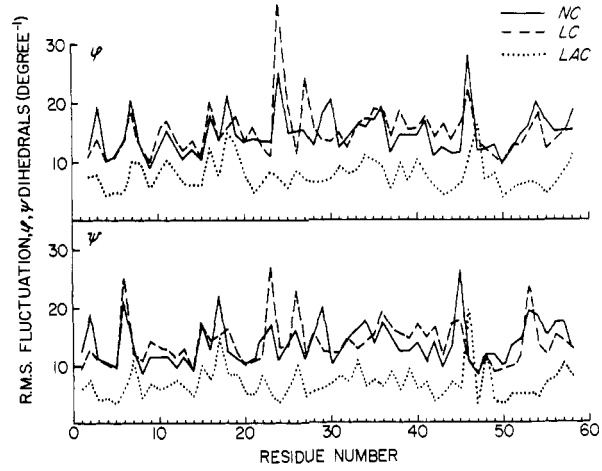

Figure 7. Root-mean-square fluctuations of the dihedrals angles ϕ,ψ over 25 "ps": (—) NC run; (---) LC run; (···) LAC run.

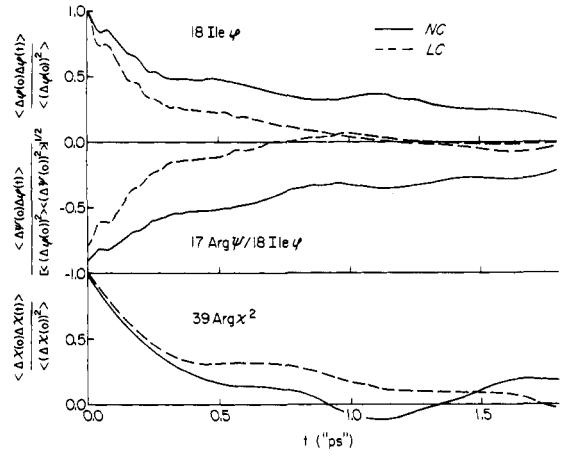

Figure 8. Auto- and cross-correlation function for backbone and side-chain dihedral angles over 25 "ps": (—) NC run; (---) LC run.

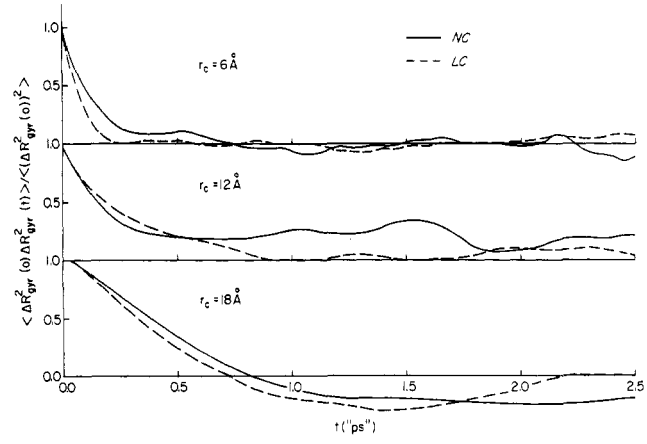

Figure 9. Radius of gyration autocorrelation function with different spherical cutoff values: (—) NC run; (---) LC run.

or the particle number density ρ, defined as the number of atoms within a sphere of radius r_c around the center of mass divided by the volume of this sphere, as a function of the cutoff radius r_c and as a function of time (see Table XX). For r_c = 6 Å the correlation in R_{gyr}^2 decays within 0.2–0.5 "ps" and no long-time correlation is observed (Figure 9). For larger r_c, the decay is slower and a correlation with a period of about 4–5 "ps" seems to exist. The core of the protein seems to be rather stable, but the surface shows a kind of oscillatory motion with a period greater than 4 "ps". It would take longer runs to obtain

Table XVIII
Flexible Backbone and Disulfide Dihedral Angles[a]

	NC						LC					
	⟨q⟩					$q_{max} - q_{min}$	⟨q⟩					$q_{max} - q_{min}$
dihedral q	0-5	5-10	10-15	15-20	20-25	0-25	0-5	5-10	10-15	15-20	20-25	0-25
6 Leu ψ	8.4	6.1	−7.4	7.4	10.8	127.6	66.8	65.4	73.7	108.4	113.7	107.1
7 Glu φ	−88.4	−90.4	−75.3	−84.3	−88.1	136.4	−136.1	−139.5	−139.2	−161.4	−165.5	99.0
15 Lys ψ	100.1	107.4	108.0	106.5	105.3	102.3	107.1	106.6	101.7	102.3	105.2	111.2
16 Ala φ	−138.2	−145.3	−147.1	−145.9	−141.7	98.0	−150.6	−141.5	−134.8	−138.2	−145.6	112.4
17 Arg ψ	83.4	87.9	101.2	100.5	75.7	112.6	107.5	93.5	96.8	106.9	101.3	104.3
18 Ile φ	−108.1	−108.4	−121.0	−120.1	−94.9	117.3	−115.5	−110.6	−120.7	−125.2	−117.8	104.2
23 Tyr ψ	99.7	90.1	99.3	91.9	82.7	99.3	82.9	71.0	79.7	118.1	123.8	134.3
24 Asn φ	−4.6	5.6	−6.4	6.5	19.8	132.6	8.1	24.2	13.0	−39.1	−49.3	171.6
26 Lys ψ	−24.3	−28.3	−29.4	−35.0	−30.7	106.9	54.3	66.4	36.3	47.5	49.0	135.0
27 Ala φ	−132.8	−140.0	−138.7	−126.3	−124.1	91.0	−135.5	−145.7	−113.8	−131.4	−133.7	144.3
29 Leu ψ	145.8	141.3	166.3	157.4	142.8	105.4	124.7	125.3	125.6	119.4	117.6	87.6
30 Cys φ	−66.0	−71.3	−92.0	−76.5	−65.8	122.7	−70.6	−77.4	−86.6	−81.6	−75.3	100.7
45 Phe ψ	112.5	162.8	160.4	166.5	167.0	159.0	152.9	168.4	159.4	156.2	163.5	104.0
46 Lys φ	−50.7	−96.7	−93.3	−104.5	−107.5	188.2	−91.6	−106.5	−97.8	−88.9	−101.2	138.0
53 Arg ψ	−48.9	−35.1	−37.3	−45.0	−23.2	111.0	16.8	7.3	9.6	14.8	22.0	122.0
54 Thr φ	−79.4	−94.5	−95.8	−87.4	−94.5	109.5	−119.6	−111.8	−117.1	−119.0	−125.0	110.6
30 Cys χ^1	−67.3	−67.9	−72.8	−68.9	−66.2	65.8	−72.6	−72.2	−118.6	−123.3	−123.3	123.3
χ^2	−63.0	−64.5	−61.3	−61.7	−67.2	85.8	−78.8	−83.3	−137.6	−143.8	−148.3	123.0
χ^{S-S}	−108.1	−104.4	−103.5	−107.3	−105.2	55.2	−106.9	−104.5	−86.9	−87.3	−90.9	71.6
51 Cys χ^1	−171.4	−177.9	−179.9	−177.2	−176.3	75.8	−159.1	−155.4	−89.9	−79.7	−79.4	140.6
χ^2	−120.4	−113.0	−111.2	−117.3	−112.3	95.4	−116.7	−116.2	−100.7	−102.5	−101.8	76.5

[a] The time-averaged value ⟨q⟩ of dihedral q is given for five consecutive periods of 5 "ps". The difference between the largest and smallest q value in 25 "ps" is denoted by $q_{max} - q_{min}$. All values are given in degrees.

Table XIX
Dihedrals Showing More Than Four Transitions[a]

	NC					LC				
dihedral q	$\langle(\Delta q)^2\rangle^{1/2}$	$q_{max} - q_{min}$	n_{tr}	Δt_{min}	Δt_{max}	$\langle(\Delta q)^2\rangle^{1/2}$	$q_{max} - q_{min}$	n_{tr}	Δt_{min}	Δt_{max}
15 Lys χ^3	25.3	129.6	0			27.7	136.2	5	0.9	8.9
χ^4	36.8	160.0	5	0.4	15.1	38.1	186.2	10	0.3	8.6
17 Arg χ^4	39.3	193.5	7	1.8	5.8	38.5	189.2	4	0.7	12.4
26 Lys χ^4	27.5	156.6	7	0.2	9.9	21.8	196.4	0		
29 Leu χ^1	46.0	202.6	5	1.2	9.2	96.4	475.2	3	2.9	14.8
χ^2	37.6	173.3	11	0.6	6.3	70.2	312.8	8	0.6	11.2
39 Arg χ^1	34.7	170.7	4	0.6	15.4	37.4	160.6	9	0.5	7.0
χ^3	36.0	200.0	4	0.6	14.3	63.6	283.7	8	0.8	8.0
χ^4	40.6	224.0	6	0.9	5.7	54.1	281.4	7	1.3	6.8
41 Lys χ^4	28.3	168.7	5	3.4	10.0	41.5	256.2	6	0.2	7.7
44 Asn χ^2	23.2	127.2	0			48.3	222.0	5	1.2	13.6
46 Lys χ^4	88.9	409.5	5	1.3	10.7	57.4	308.7	10	0.2	9.3
52 Met χ^3	53.1	225.4	6	1.9	7.4	61.4	258.0	5	0.5	11.4
53 Arg χ^4	47.7	252.7	5	2.0	6.8	36.4	187.2	5	1.1	8.2

[a] The root-mean-square fluctuation $\langle(\Delta q)^2\rangle^{1/2}$ and the difference $q_{max} - q_{min}$ between the largest and smallest value of dihedral q is computed over 25 "ps". The number of transitions in this period is denoted by n_{tr}. The shortest time between two observed transitions is denoted by Δt_{min} and the largest one by Δt_{max}.

statistically valid data concerned with these slow motions. Relative to the X-ray structure, the molecule is somewhat too tightly packed, especially the core, which is explained by the vacuum boundary condition; there is no external medium to attract the atoms of the protein. The difference between ref 11 and the present results is due to the larger cutoff radius used for the nonbonded interactions (see above).

Summarizing, we conclude that the NC run (equilibration: 30 "ps", analysis: 25 "ps") yields essentially the same results as previous MD runs of BPTI, that is, that of ref 10 (equilibration: 3.5 "ps", analysis: 9 "ps") and that of ref 11 (equilibration: 75 "ps", analysis: 100 "ps"). The small differences can be explained by the different equilibration and analysis periods and by the slightly different potentials that were employed.

5. Molecular Dynamics with Bond-Length Constraints (LC)

In this section we compare the LC run with the NC run to determine the effects of freezing the bond-length degrees of freedom. On a time scale smaller than 0.05 "ps", differences must occur since the bond-length vibrations have a period of approximately 0.03 "ps". However, we are not concerned with such short times and consider effects that occur on a time scale longer than 0.05 "ps". The comparison of processes that occur on a time scale of the order of the total time covered by the MD runs (25 "ps") is also

Table XX
Radius of Gyration and Number Density[a]

averaging period, "ps"	$R_{gyr}(\langle x\rangle)$, Å ($r_c = \infty$)			$\rho(\langle x\rangle)$, Å^{-3} (r_c = 8 Å)		
	NC	LC	LAC	NC	LC	LAC
0-5	10.18	10.11	10.12	0.0704	0.0695	0.0709
5-10	10.18	10.11	10.15	0.0704	0.0709	0.0709
10-15	10.16	10.12	10.16	0.0709	0.0690	0.0695
15-20	10.16	10.12		0.0704	0.0634	
20-25	10.19	10.11		0.0704	0.0625	
0-25	10.17	10.11	10.14	0.0709	0.0685	0.0709
X-ray	10.96			0.0597		
ref 12	10.22			0.063		

[a] The values of the radius of gyration R_{gyr} and the number density ρ are calculated from the various time-averaged structures denoted by $\langle x\rangle$. Only atoms that are within r_c from the center of mass are taken into account.

not meaningful, because of insufficient statistical accuracy. Moreover, the comparison of the 25-"ps" averages of the two runs may be distorted by possible effects on the averages of slow processes (time scale >5 "ps"). Such effects can be detected and eliminated by comparing the averages of the 5-"ps" averages for the two runs. Thus, it is appropriate to examine similarities and differences of the NC and LC runs that involve times from 0.05 to 5 "ps".

The temperatures of the two runs are slightly different (Table X). The LC run shows larger fluctuations in the temperature because of its lower number of degrees of freedom. The calculated difference corresponds to the reduction in the number of degrees of freedom; that is, $10.6/8.6 = (1368/900)^{1/2}$. The heat capacities differ, when corrected for the difference in number of degrees of freedom, by about 5%. This is within the accuracy that is expected for a quantity that depends on fluctuations in the ensemble. The kinetic energy is equally distributed over the backbone (Figure 1) and no particularly hot or cold side-chain atoms are observed (Table XI). When the average temperature for the atom types, $\overline{\langle T\rangle}$, is compared, the largest differences are observed for the C_α and the C_{ϵ_1} atoms. For the latter, this can be understood from the small (8) number of atoms present in the molecular average. For the C_α atoms, this is not a sufficient explanation. In the NC run the C_α atoms have an average covalent bond energy of 0.492 kcal mol^{-1}. If we assume that the part of the $\overline{\langle E_{kin}\rangle}$ alone the bond-length degrees of freedom is equal to the bond energy (harmonic oscillator), we find $\overline{\langle E_{kin}\rangle} = 0.412$ for the remaining degrees of freedom of the C_α atoms, yielding $\overline{\langle T\rangle} = 269$ K for these degrees of freedom, to be compared to $\overline{\langle T\rangle} = 287$ K for the LC run. This difference is in the range of temperature fluctuations for different atoms found in the same run (e.g., N and C_α in the LC run).

The velocity autocorrelation functions for the C_α atoms show the same basic features with and without constraints when the high-frequency ($\nu \gtrsim 25$ "ps"$^{-1}$) oscillations have been eliminated (Figure 2). The dominant peaks in the spectral density occur at $\nu_1 = 14$ "ps"$^{-1}$ and $\nu_2 = 19$ "ps"$^{-1}$ (NC: $\nu_1 = 13$ "ps"$^{-1}$, $\nu_2 = 17$ "ps"$^{-1}$).

The root-mean-square deviation of the time-averaged structure from the X-ray structure amounts to 3.0 Å for all atoms (NC: 2.7 Å) and 2.2 Å for the C_α atoms (NC: 2.1 Å). For the difference between the NC and the LC averaged structures, the corresponding values are 1.2 Å for all atoms and 0.9 Å for the C_α atoms. Thus, the MD structures are closer to each other than to the X-ray structure. The C_α atoms that show a NC/LC root-mean-square difference larger than 0.9 Å are at the terminals of the chain (residues 1–6, 57, and 58) and in the extended loop (residues 25–29). These are the parts of the protein that show the largest root-mean-square fluctuations (Figure 3). Apart from the region around 5 Cys, the pattern of the C_α root-mean-square fluctuations is similar to that of the NC run, but shifted to larger values. The LC root-mean-square fluctuations are 20% larger than the NC fluctuations for the C_α atoms and for all atoms in the protein (Table XII). This is due to a slow (time scale >5 "ps") motion that occurs in the LC run, since the 5-"ps" root-mean-square fluctuations differ by less than 4%. This is confirmed by the agreement of the 5-"ps" C_α root-mean-square fluctuation curves in Figure 4. The slow mode in the LC run does not show up very clearly in the fluctuations of atoms at the surface of the protein (Lys N_ζ, Table XIII), since such fluctuations are determined mainly by local motions. The largest fluctuations (1.2–2.9 Å) are found in residues 52 Met, 39 Arg, 31 Gln, 29 Leu, 44 Asn, 17 Arg, 10 Tyr, 6 Leu, and 26 Lys. Except for the N_{δ_2} atom in 44 Asn, the same atoms show large fluctuations in the NC run. The atom fluctuation anisotropies are sensitive to the slow mode occurring in the LC run. Therefore, only the 5-"ps" values for the NC run and LC run should be compared. These are fairly close to each other and follow the same pattern as a function of atom name (Table XIII). The positional fluctuation autocorrelation functions decay similarly with and without constraints, except for possible long-time (>5 "ps") tails (Figure 5), which represent slow motions that occur in one run but not in the other. The NC and LC relaxation times are of the same order of magnitude, except for the atoms in 26 Lys (Table XIV).

The time-averaged bond angles and improper dihedral angles of the LC run are very close to those of the NC run; that is, their root-mean-square difference over all angles in the protein is about 1° (Table XV). Also the fluctuations of these quantitites are very similar; for example, the root-mean-square difference between the NC and LC root-mean-square fluctuations of the (N–C_α–C) bond angles equals 0.3°. The LC bond-angle autocorrelation functions exhibit the same features as those of the NC bond angles (Figure 6). The spectral density shows the same pattern in both cases. This indicates that the bond-length degrees of freedom are only weakly coupled to the bond angles. The same picture emerges for the peptide dihedral angles ω (Figure 6). The time-averaged backbone dihedral angles for the NC and LC runs differ generally by less than 14°, except for the dihedral angles at the top and the bottom of the protein and in the extended loop. Again the NC and LC averages are closer to each other than to the X-ray values. The fluctuations of the backbone dihedral angles for the two runs are remarkably close (Table XVI). When side-chain dihedrals are included, the differences are larger. The LC root-mean-square fluctuations are systematically slightly larger than the NC fluctuations, which is due to the slow mode occurring in the LC run. Even the root-mean-square fluctuations of the individual ϕ,ψ angles are much the same (Figure 7). The differences between the NC and LC curves in Figure 7 arise from rare transitions (time scale $\gtrsim$5 "ps") between two favorable angular conformations (see section 4), as is made clear by the examples shown in Table XVIII. The correlation between the motions of neighboring bond angles or dihedral angles is generally the same with and without constraints (Table XVII). Also the dihedral autocorrelation and cross-correlation functions look the same, apart from their long-time tails, which may be different when infrequent transitions occur in one run and not in the other (Figure 8). For example, the 17 Arg ψ,18 Ile ϕ

dihedral angles show a slow transition in the NC run (Table XVIII, Figure 7) that does not occur in the LC run and caused different long-time decays in Figure 8. During the LC run, 119 dihedral transitions were observed, none of them in the backbone (NC: 110 total and 3 in the backbone). A transition in the 51 Cys χ^1 dihedral occurred after 11 "ps", which was accompanied by a rearrangement of the disulfide bridge (Table XVIII). The sum of the 5-"ps" time averages of the five side-chain dihedral angles forming the 30 Cys–51 Cys disulfide bridge differs only 2° before and after the rearrangement, as can be observed from Table XVIII. Generally, the dihedral angles that exhibit the largest number of transitions are the same for both runs (Table XIX).

The dynamical behavior of the radius of gyration R_{gyr} is very similar with and without constraints (Figure 9) and so are the time-average values (Table XX). The variation in the number density (5-"ps" values) suggests that a slow global mode occurs in the LC run and is not present in the NC run.

Summarizing, we find that no significant differences on a time scale of 0.05–5 "ps" were observed between the MD run without constraints (NC) and that with bond-length constraints (LC). On a time scale shorter than 0.05 "ps", the NC run exhibits high-frequency motions due to the bond-length vibrations that are not present in the LC run. On a time scale comparable to the time span covered by the MD runs (25 "ps") the NC run differed somewhat from the LC run. The latter showed a slow, more or less collective motion (time scale >5 "ps") of the protein. Whether this slow mode is a real difference between the NC and the LC runs or is due to poor statistics can be decided only by performing MD runs over much longer periods ($\gtrsim$100 "ps") than the characteristic time of the slow mode. However, since freezing the bond lengths has only slight effects on the dynamics of the protein on a time scale of 0.05–5 "ps", it is not likely that significant differences will occur on longer time scales.

6. Molecular Dynamics with Bond-Length and Bond-Angle Constraints (LAC)

In this section we compare the LAC run with the other (NC and LC) runs to determine the effects of freezing both the bond-length and bond-angle degrees of freedom. The comparison will be much less detailed than that in the former section, since it is clear from the examination of a few quantities that the motion of the protein is severely limited by freezing the bond-angle degrees of freedom.

The LAC run shows larger temperature fluctuations (12.2 K) than the other runs because of its lower number of degrees of freedom (Table X). After correcting for the reduction in the number of degrees of freedom, we find the value $(278/1368)^{1/2} \times 12.2 = 5.5$ K, which is much smaller that the values of the NC and LC runs (8.6 K). This effect of rigid bond angles is also reflected in the heat capacity of 1.96 k, which is closer to the hard-sphere value of 1.5 k than to the harmonic value of 3.0 k. Since it is not evident how to assign the remaining degrees of freedom to the individual atoms in the LAC case, no average temperatures per atom name are given in Table XI.

The root-mean-square deviation of the time-averaged structure from the X-ray structure is 2.8 Å for all atoms (NC: 2.7 Å, LC: 3.0 Å) and 2.3 Å for the C_α atoms (NC: 2.1 Å, LC: 2.2 Å). The corresponding quantities for the NC/LAC difference are 0.5 and 0.3 Å and for the LC/LAC difference 1.3 and 1.0 Å. We note that both the LC run and the LAC run were started from the NC configuration that had been obtained after 20 "ps" of equilibration (section 3). The fluctuations in the atom positions are quenched by a factor of 2 in LAC (Table XII). Even side-chain atoms in residues at the surface of the protein show the reduction of their mobility by a factor of 2 (Table XIII). Only three atoms showed root-mean-square positional fluctuations larger than 1 Å: 17 Arg N_{η_2} (1.1 Å), 52 Met S_δ (1.0 Å), and C_ϵ (1.3 Å). By contrast, in the NC run 25 atoms showed fluctuations larger than 1 Å. However, the pattern of the C_α atom root-mean-square fluctuations as a function of residue number still exhibits the same qualitative features as that of the NC run (Figures 3 and 4). The major differences are the greatly reduced flexibility of the extended loop (residues 25–29) and the slightly diminished flexibility near the binding site (residues 17–19). The similarity of the curves in Figures 3 and 4 suggests that the pattern is determined mainly by the way the chain is folded and the resulting nonbonded contacts[12,27] and that these are mainly preserved even though there are small differences in the average structures. The LAC 5-"ps" fluctuation anisotropies are further from the NC values than from those of the LC run (Table XIII). A comparison of the time-averaged dihedral angles confirms that the LAC structure is very close to the NC structure (Table XVI). The dihedral fluctuations are strongly inhibited by freezing the bond angles (Table XVI and Figure 7). Only eight dihedrals show root-mean-square fluctuations larger than 20° (NC: 51, LC: 59) and only four transitions are observed during 15 "ps" (NC: 110, LC: 119 in 25 "ps"); that is, two transitions (120° → −120° → 120°) for 39 Arg χ^4 and two transitions (−90 → −270° → −90°) of 52 Met χ^3. Thus, the dihedral angle motion is severely inhibited by the bond-angle constraints and the resulting rigidity of the molecule.

The radius of gyration R_{gyr} and the number density ρ are the same as for the NC run (Table XX).

Summarizing, we conclude that the dynamics of the protein is strongly altered by freezing the bond-angle degrees of freedom. Because of the close-packed nature of the native protein, in which torsional barriers are dominated by nonbonded contacts, the rigidity imposed on the chain by the application of bond-length and bond-angle constraints makes it difficult for the atoms to pass by each other; i.e., the effect of the nonbonded repulsions ("excluded volume") in inhibiting the motion is significantly increased. The dynamical fluctuations are quenched by a factor of 2 and the dihedral transition rates are dramatically reduced. Nevertheless, the pattern of relative fluctuations as a function of residue number remains similar with bond-angle constraints to that obtained in their absence.

7. Concluding Discussion

The molecular dynamics (MD) simulation of the protein BPTI that has been performed without applying any constraints (NC) yields essentially the same results as previous studies for this molecule.[10,11] The small differences between the various runs are explained by the different lengths of the equilibration and analysis periods and the minor differences in the potential functions. The correspondence of the results from the series of runs that have now been made supports the validity of molecular dynamics simulations of proteins. The length of the analysis period determines which phenomena can be studied. Phenomena or motions that occur on a time scale equal to or longer than the analysis period may distort the time averages taken over the whole analysis period. These slow motions can be detected and eliminated by performing time averages over separate parts of the analysis period and then averaging (ensemble average over the partial time averages). This procedure may be very useful for deter-

mining the sensitivity of the MD results to the time span covered by the run and to obtain improved averages for quantities that involve shorter time fluctuations.

When bond-length constraints are applied to the molecule (LC), no significant differences relative to the unconstrained MD run (NC) are observed for most of molecular properties studied; only the average structure is somewhat different in the constrained run (~1.2 Å root-mean-square for all atoms); this is not necessarily due to the constraints, per se, but may simply be a consequence of the difference in initial conditions (particularly the velocities), which leave the equilibrated system in a somewhat different region of conformation space. Although it is impossible to prove rigorously that the constrained and unconstrained models yield the same dynamics, the results obtained justify the conclusion that freezing the bond lengths does not significantly alter the dynamics of the protein on a time scale greater than 0.05 "ps". The bond-length degrees of freedom appear to be weakly coupled to the other degrees of freedom of the molecule. This means that for the investigation of many questions the efficiency of computer simulations of macromolecules can be improved by about a factor of 3 by the application of bond-length constraints.

When bond-length and bond-angle constraints are applied (LAC), the dynamics of the molecule is severely altered. The positional and dihedral angle fluctuations are quenched by about a factor of 2, although the relative magnitudes of the fluctuations in different parts of the molecule are approximately preserved. Also, the dihedral angle transition rates are dramatically reduced. The bond-angle degrees of freedom appear to play an important role in the dynamics of the protein; that is, they are strongly coupled to the other degrees of freedom of the molecule. Although the fluctuations in the bond angles are small, they are essential for the motion of the molecule, due to the importance of nonbonded interactions in determining the magnitudes of the fluctuations. This means that macromolecular models that are based on the assumption of rigid (frozen) bond angles yield a physically incorrect picture of the molecular dynamics; it may be possible to introduce modified dihedral angle and other potential terms to account approximately for the bond-angle relaxation. Similar conclusions have been obtained from constrained energy minimization studies of macromolecules.[24]

A number of recent theoretical studies of alkane polymer chains have addressed the question of the effect on the chain motions of coupling between different internal degrees of freedom.[21,48-51]. It is of considerable interest to compare the results of these studies for more extended systems with those found here for a densely packed protein. Closest to the present work is an examination[48,49] of dihedral angle transitions of polymer chains by a multidimensional extension of the Kramers formulation. It was found that a localized reaction coordinate can be constructed, which in addition to the dihedral angle of interest involves small contributions from neighboring dihedral angles and the adjacent bond angles; the bond-length contribution appears to be negligible. These conclusions concerning the form of the reaction coordinate are confirmed by simulation studies in which the bond-angle and bond-length force constants were varied. Thus, the analyses of isomerization in hydrocarbon chains are in accord with the protein results in that the rate of dihedral angle transitions for amino acid side chains is greatly reduced by bond-angle constraints and not significantly altered by bond-length constraints. The importance of constraints on excluded volume effects that reduce the magnitude of the root-mean-square atomic fluctuations in the protein simulation has not been examined in polymers.

Acknowledgment. We gratefully acknowledge many stimulating discussions with B. Olafson and S. Swaminathan.

References and Notes

(1) Ryckaert, J.-P.; Bellemans, A. *Chem. Phys. Lett.* **1975**, *30*, 123.
(2) Ryckaert, J.-P.; Bellemans, A. *Faraday Discuss. Chem. Soc.* **1978**, *66*, 95.
(3) Weber, T. A. *J. Chem. Phys.* **1978**, *69*, 2347.
(4) Bishop, M.; Kalos, M. H.; Frisch, H. L. *J. Chem. Phys.* **1979**, *70*, 1299.
(5) Rebertus, D. W.; Berne, B. J.; Chandler, D. *J. Chem. Phys.* **1979**, *70*, 3395.
(6) Weber, T. A. *J. Chem. Phys.* **1979**, *70*, 4277.
(7) Weber, T. A.; Helfand, E. *J. Chem. Phys.* **1979**, *71*, 4760.
(8) Rossky, P. J.; Karplus, M.; Rahman, A. *Biopolymers* **1979**, *18*, 825.
(9) Rossky, P. J.; Karplus, M. *J. Am. Chem. Soc.* **1979**, *101*, 1913.
(10) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature (London)* **1977**, *267*, 585.
(11) Karplus, M.; McCammon, J. A. *Nature (London)* **1979**, *277*, 578.
(12) McCammon, J. A.; Wolynes, P. G.; Karplus, M. *Biochemistry* **1979**, *18*, 92.
(13) Northrup, S. H.; Pear, M. R.; McCammon, J. A.; Karplus, M.; Takano, T. *Nature (London)* **1980**, *287*, 659.
(14) McCammon, J. A.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 3585.
(15) Fixman, M. *J. Chem. Phys.* **1978**, *69*, 1527, 1538.
(16) Ceperley, D.; Kalos, M. H.; Lebowitz, J. L. *Phys. Rev. Lett.* **1978**, *41*, 313.
(17) Helfand, E.; Wasserman, Z. R.; Weber, T. A. *J. Chem. Phys.* **1979**, *70*, 2016.
(18) Levy, R. M.; Karplus, M.; McCammon, J. A. *Chem. Phys. Lett.* **1979**, *65*, 4.
(19) Pear, M. R.; Weiner, J. H. *J. Chem. Phys.* **1979**, *71*, 212.
(20) van Gunsteren, W. F.; Berendsen, H. J. C.; Rullmann, J. A. C. *Mol. Phys.* **1981**, *44*, 69.
(21) Montogomery, J. A., Jr.; Holmgren, S. C.; Chandler, D. *J. Chem. Phys.* **1980**, *73*, 3688.
(22) Flory, P. J. "Statistical Mechanics of Chain Molecules"; Interscience: New York, 1969.
(23) Scheraga, H. A. *Chem. Rev.* **1971**, *71*, 1975.
(24) van Gunsteren, W. F.; Karplus, M. *J. Comput. Chem.* **1980**, *1*, 266.
(25) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(26) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.
(27) Gelin, B. R.; Karplus, M. *Biochemistry* **1979**, *18*, 1256. Gelin, B. R.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 801.
(28) Gibson, K. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 420.
(29) Ramachandran, G. N.; Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 283.
(30) Scheraga, H. A. *Adv. Phys. Org. Chem.* **1968**, *6*, 103.
(31) Levitt, M.; Lifson, S. *J. Mol. Biol.* **1969**, *46*, 269.
(32) Ermer, O.; Lifson, S. *J. Mol. Spectrosc.* **1974**, *51*, 261.
(33) Levitt, M. *J. Mol. Biol.* **1974**, *82*, 393.
(34) IUPAC–IUB Commission on Biochemical Nomenclature *Biochemistry* **1970**, *9*, 3471.
(35) Margenau, H.; Kestner, N. R. "Theory of Intermolecular Forces"; Pergamon Press: New York, 1969.
(36) Taylor, B. N.; Parker, W. H.; Langenberg, D. N. *Rev. Mod. Phys.* **1969**, *41*, 375.
(37) Brookhaven Protein Data Bank, Brookhaven National Laboratory, Upton, N.Y., 1978. Deisenhofer, J. O.; Steigemann, W. R. *Acta Crystallogr., Sect. B* **1975**, *B31*, 238.
(38) Verlet, L. *Phys. Rev.* **1967**, *159*, 98.
(39) van Gunsteren, W. F.; Berendsen, H. J. C.; Rullmann, J. A. C. *Faraday Discuss. Chem. Soc.* **1978**, *66*, 58.
(40) Levy, R. M., unpublished calculations.
(41) van Gunsteren, W. F. *Mol. Phys.* **1980**, *40*, 1015.
(42) Lebowitz, J. L.; Percus, J. K.; Verlet, L. *Phys. Rev.* **1967**, *153*, 250.
(43) van Gunsteren, W. F.; Karplus, M. *Biochemistry*, in press.
(44) McLachlan, A. D. *J. Mol. Biol.* **1979**, *128*, 49.
(45) McCammon, J. A.; Karplus, M. *CRC Crit. Rev. Biochem.* **1981**, *4*, 293.

(46) Frauenfelder, H.; Petsko, G. A.; Tsernoglou, D. *Nature (London)* **1979**, *280*, 558.
(47) Artymiuk, P. J.; Blake, C. C. F.; Grace, D. E. P.; Oatley, S. J.; Phillips, D. C.; Sternberg, M. J. E. *Nature (London)* **1979**, *280*, 563.
(48) Skolnick, J.; Helfand, E. *J. Chem. Phys.* **1980**, *72*, 5489.
(49) Helfand, E.; Wasserman, Z. R.; Weber, T. A.; Skolnick, J.; Runnels, J. H. *J. Chem. Phys.* **1981**, *75*, 4441.
(50) Evans, G. T.; Knauss, D. C. *J. Chem. Phys.* **1980**, *72*, 1504.
(51) Knauss, D. C.; Evans, G. T. *J. Chem. Phys.* **1981**, *74*, 4621.

Hydrodynamic Properties of a Torsionally Flexible Cylinder: An Application to DNA

Stuart A. Allison

Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403.
Received February 16, 1982

ABSTRACT: The hydrodynamic properties of a torsionally flexible cylinder are studied using the extrapolated shell method. The method appears to be exact. What is derived is a general equation that relates the angular velocity at some point along the cylinder to the torque per unit length acting on the entire cylinder. In the special case of a torque that varies sinusoidally along the cylinder, a simple equation for the rotational friction coefficient is derived. This result is important in connection with the dynamical theory of torsion in macromolecules in which only approximate account of hydrodynamic interaction has been taken to date. In the case of DNA, it is concluded that present theory is valid down to approximately 5 ns and that by taking hydrodynamic interaction properly into account, theory would be valid down to about 0.1 ns.

Introduction

In this article, the hydrodynamic properties of a long cylinder that is torsionally flexible are considered. Specifically, a general equation is derived relating the instantaneous angular velocity at some point to the instantaneous torque per unit length acting over the entire cylinder. This result appears to be exact within the limitations of the assumptions about the nature of the fluid and its interaction with the cylinder surface. In the special case of a long cylinder undergoing uniform rotation about the cylindrical axis, Perrin's result is obtained[1]

$$\tau_{\text{hydro}} = \gamma_{\text{p}}\omega \tag{1}$$

$$\gamma_{\text{p}} = 4\pi\eta a^2 \tag{2}$$

where τ_{hydro} is the hydrodynamic torque per unit length, ω is the angular velocity, γ_{p} is the rotational friction coefficient per unit length, η is the solvent viscosity, and a is the radius of the cylinder.

This work was motivated by the recent interest in the rapid twisting or torsional motions of DNA detected by fluorescence depolarization[2-4] and EPR.[5] The dynamical theory of DNA twisting was originally worked out independently by Barkley and Zimm[6] and Allison and Schurr.[7] Since then, the effect of torsional motions on NMR relaxation parameters[8] and depolarized light scattering[9] have been studied. It can be anticipated that torsional motions in macromolecules will be the subject of numerous future investigations.

To date, the hydrodynamic interactions of a torsionally flexible macromolecule have only been accounted for in an approximate way. Both Barkley and Zimm[6] and Allison and Schurr[7] used the Perrin formula (eq 1) to account for the hydrodynamic torque. In any dynamical model, this is exact only if the entire molecule (modeled as a cylinder) rotates at the same rate. If, for example, a small region of the cylinder rotates faster than neighboring regions at a given instant, then the hydrodynamic torque on the faster moving region would be larger than predicted using the Perrin formula because of increased solvent drag. One of the primary objectives of this work is to account more accurately for the hydrodynamic interactions in a dynamical model of macromolecular torsion.

The problem is approached using the extrapolated shell method.[10,11] This method originated with Bloomfield et al.[12] and consists basically of viewing the hydrodynamic properties of the macromolecule as resulting from the solvent perturbation by point sources of friction uniformly distributed over the object's surface. The problem of the hydrodynamic interaction between point sources of friction was considered by Oseen,[13] who derived an interaction tensor, the Oseen tensor, for any configuration. This tensor has been extended to account for the finite size of the (spherical) beads and is known as the modified Oseen tensor in this form.[14,15]

The macromolecule is modeled by placing a large number of identical beads uniformly over its surface and the hydrodynamic properties (translational and rotational friction coefficients) are determined using the Oseen tensor or its modified form. In general, the problem must be solved numerically for a finite number of beads. The process is repeated using a larger number of smaller beads and finally, the limit of an infinite number of beads of infinitesimal size is estimated by extrapolation.[10,11] In this limit, the shell model appears to be exact and is of great value and utility since it can, in principle, be applied to macromolecules of any geometry. Unfortunately, these calculations are expensive and time-consuming. Garcia de la Torre and Bloomfield have described various approximations and possible alternatives to the extrapolated shell method.[16,17]

In a few limited cases where the object possesses a high degree of symmetry, it is possible to solve the problem analytically. A still broader class of problems can be solved analytically if, for example, use is made of the Kirkwood approximation.[18] Recently, the translational friction of a toroid was solved in this manner and was found to be accurate to within a few percent of the exact value.[19] (In the case of the toroid, the exact value was determined numerically by the extrapolated shell method with no approximations. These calculations required more than 200 h of computer time on a PDP-12 computer equipped

0024-9297/82/2215-1544$01.25/0