PROTEINS: Structure, Function, and Genetics 23:49–62 (1995)

Structure and Internal Dynamics of the Bovine Pancreatic Trypsin Inhibitor in Aqueous Solution From Long-Time Molecular Dynamics Simulations

R.M. Brunne,[1] K.D. Berndt,[2] P.Güntert,[2] K. Wüthrich,[2] and W.F. van Gunsteren[1]
[1]Laboratorium für Physikalische Chemie, ETH Zentrum, CH-8092 Zürich, and [2]Institut für Molekularbiologie und Biophysik, ETH Hönggerberg, CH-8093 Zürich, Switzerland

ABSTRACT **Structural and dynamic properties of bovine pancreatic trypsin inhibitor (BPTI) in aqueous solution are investigated using two molecular dynamics (MD) simulations: one of 1.4 ns length and one of 0.8 ns length in which atom–atom distance bounds derived from NMR spectroscopy are included in the potential energy function to make the trajectory satisfy these experimental data more closely. The simulated properties of BPTI are compared with crystal and solution structures of BPTI, and found to be in agreement with the available experimental data. The best agreement with experiment was obtained when atom–atom distance restraints were applied in a time-averaged manner in the simulation. The polypeptide segments found to be most flexible in the MD simulations coincide closely with those showing differences between the crystal and solution structures of BPTI.**

Key words: BPTI, structure refinement, time-averaged NOE restraints, protein dynamics

INTRODUCTION

With computers becoming increasingly powerful, simulation of dynamic processes is a rapidly growing area of research. The simulation of molecular dynamics has the attractive feature that motions on the atomic level can be monitored and analyzed in atomic detail. Experimentally, X-ray diffraction techniques yield a detailed atomic picture of a molecular system in single crystals, but it contains no information on the time scales of the atomic motions. NMR yields a detailed atomic resolution picture of molecules in solution, which can also include time scales of a wide range of dynamic processes, but its use is limited to systems with molecular weights smaller than about 30,000 Da.[1] Other spectroscopic techniques allow for the measurement of some relaxation times, but they cannot provide atomic spatial resolution. In view of these limitations of experimental measuring techniques, computer simulation methods are an alternative for obtaining a dynamic picture of biomolecular motion, and for complementing experimental data.

A necessary condition for the use of computer simulation techniques to study molecular motion is that the molecular model and the interatomic interaction function, V_{phys}, are sufficiently reliable to predict various properties of the molecular system of interest. A second necessary condition is that the simulation spans the time scale of the motions that determine the properties of interest. These two conditions limit the applicability of molecular dynamics (MD) computer simulation techniques to biomolecular systems considerably.

If the molecular model and atomic interaction function V_{phys} were perfect, and a simulation could be carried on infinitely long, the generated trajectory would exactly represent the real molecular system: the structural and dynamic properties derived from such a trajectory would give a complete and correct picture of the molecular system, without recourse to experimental data. In practice, neither condition is fulfilled. Atomic interaction functions are not infinitely accurate due to various approximations made in the derivation of V_{phys}, for example, with respect to electronic degrees of freedom, quantum-mechanical effects, many-body interactions, etc. Furthermore, a computer simulation cannot be carried out for infinitely many steps: generally, a MD simulation of a molecular system does not extend beyond the nanosecond time scale, which is not sufficiently long to sample the entire range of motions of a protein in solution.

One way to improve the reliability of a molecular simulation is to incorporate experimental information on the specific molecular system of interest in

Abbreviations used: BPTI, bovine pancreatic trypsin inhibitor; MD, molecular dynamics; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; rms, root mean square; rmsd, root mean square displacement.

Received March 17, 1995; accepted April 14, 1995.
Address reprint requests to W.F. van Gunsteren, Laboratorium für Physikalische Chemie, ETH Zentrum, CH-8092 Zürich, Switzerland.
Present address of R. M. Brunne, Bayer AG, Pharma Forschungszentrum, Aprather Weg, D-42096 Wuppertal, Germany.

the simulation. This can be done by adding to the standard physical potential energy function V_{phys} an extra term V_{restr} which restrains or influences the motion such that the generated trajectory yields average properties in accordance with the experimental information on the specific molecular system. For example, when distance (upper) bounds r_{ij}^{ub} have been derived from nuclear Overhauser effect (NOE) cross-peak intensities originating from nuclei i and j, the function[2]

$$V_{restr} = \frac{1}{2} k_{dr} \sum_{\text{NOE pairs } (i,j)} [max(0, \langle r_{ij}^{-p} \rangle \bar{\tau}_{dr}^{1/p} - r_{ij}^{ub})]^2 \quad (1)$$

can be used to make the trajectory satisfy the NOE data. The function *max* delivers the largest of its two arguments, and the distance between atoms i and j is denoted by r_{ij}. The distance dependence of the dipolar interaction leads to the value $p=6$ for the exponent in (1). However, if the averaging period τ_{dr} in the time average $\langle \cdots \rangle_{\tau_{dr}}$ corresponds to a short time scale, as is usually the case in biomolecular simulations, the influence of angular fluctuations should be neglected,[3] which leads to a value of $p=3$.

Generally speaking, a MD simulation of a protein may serve different purposes.[4,5] First, the simulated molecular properties may be compared to measured ones in order to validate the atomic interaction function V_{phys} and computational procedure used in the simulation. Second, a MD simulation in which restraints, such as in (1), representing measured data are used, may be used to generate a trajectory of protein structures that satisfies the measured data. Third, such a MD trajectory may be used to analyze molecular or atomic properties that are inaccessible to measurement, from which analysis insight into protein properties may be obtained. The analysis of the MD simulations of the small protein bovine pancreatic trypsin inhibitor (BPTI) presented below aims at serving all of these purposes: (1) test of the force field used (GROMOS[6]), (2) generation of a trajectory of BPTI conformations that is compatible with the NOE data available for this protein,[7] and (3) analysis of a number of dynamic and structural properties of BPTI using the generated trajectories.

Its small size (58 residues), high stability and ready availability make BPTI suitable for testing new experimental and theoretical methods. Three different crystal structure forms have been determined from X-ray and neutron diffraction,[8–10] and were compared by Wlodawer et al.[10,11] Many structural and dynamic properties were analysed by means of NMR spectroscopy.[7,12–17] The first simulation of BPTI in vacuo was done for 20 ps by McCammon et al.[18] The first simulations of BPTI in a Lennard–Jones fluid[19] for 25 psec, in water[20] for 20 ps, and in the crystalline state[21] for 20 ps were really not sufficiently long to allow for a reliable description of structural and dynamic properties. Levitt and Sharon[22] and Daggett and Levitt[23] reported simulations of BPTI in water of 210 and 550 ps lengths, respectively. In the latter study, simulations were performed at different temperatures in order to investigate the thermal unfolding of BPTI in atomic detail. The focus of the present BPTI study is different. First, the 800 and 1400 ps MD simulations of BPTI in water are used to validate the GROMOS force field by comparison with NMR and X-ray data. Second, the effect of different ways to account for experimental NOE information in a MD simulation by the application of instantaneous ($\tau_{dr} = 0$) or time-averaged ($\tau_{dr} > 0$) atom-atom distance restraints (1) is evaluated. Third, structural and dynamic features of BPTI are analyzed.

We will present the time dependence of some global properties such as the potential energy, Lennard–Jones and Coulomb energies, the radius of gyration, and average atomic positional displacements, which are generally used to judge the quality of equilibration and stability of a simulation. We will further analyze backbone dihedral angle transitions, and average the coordinates of BPTI over time intervals between the observed transitions. These averaged conformations will then be compared to each other and to the X-ray and NMR structures. It is not our aim to characterize the conformations of BPTI completely, because that would require an analysis of the side chain dihedral angle transitions as well, which is beyond the scope of this study. Averaging the coordinates over periods during which the backbone dihedral angles do not show large changes will provide meaningful representative structures that are more suited for comparisons than an arbitrarily picked conformation at a single time point or a structure which is an average over the complete simulation. The comparison between simulated and X-ray crystal structures will focus on those regions for which the three crystal structures differ from each other, in order to investigate whether structural variation in different crystal structures is a reliable indication of increased mobility in solution as well. A detailed comparison of the NMR solution structure with the crystal structures has been given by Berndt et al.[7] A variety of other data from the BPTI simulations presented here have already been analyzed and published elsewhere: hydration properties,[24] dielectric properties,[25] and the translational and rotational diffusion properties.[26]

COMPUTATIONAL METHODS

The simulations were performed using the GROMOS force field.[6] Nonpolar hydrogen atoms were included with the carbon atoms (united atom approach), and polar hydrogen atoms were treated explicitly. In accordance with the experimental conditions chosen[27,28] for the NMR investigations of the hydration of BPTI in aqueous solution, we applied a simulation temperature of 277 K and used proto-

nated residues Arg^+ and Lys^+. We also chose the acidic residues Glu and Asp to be charged, because of experimental evidence[29] that even Asp ($pK_a \approx 3.2$) is at least 50% deprotonated at pH 3.5. The total charge of BPTI is then +6 *e*. For the water molecules the SPC/E model[30] was used. At 300 K it yields a diffusion constant that is close to the experimental value, and it also yields a resolved second peak in the O-O radial distribution function which indicates good structural properties.

As starting structure, we used a conformer from a preliminary distance geometry calculation[7] based on NMR data using the program DIANA. A least-squares fit of all backbone atoms of this NMR conformer to those of the crystal structure II[9] using conformation A of the two possible conformations of Glu-7 and Met-52 was performed (rmsd = 0.136 nm). Then, the coordinates of the four internal water molecules of the X-ray structure (W111, W112, W113, W122 from the Brookhaven data bank entry 5PTI) were added as water molecules W1, W2, W3, W4 to the distance geometry structure. Fifty steps of energy minimization with the steepest descent method were applied. The resulting structure was placed in a truncated octahedron of equilibrated SPC/E water. A distance of 0.85 nm between the protein and the square planes of the box edges and a minimum distance of 0.23 nm between the nonhydrogen protein atoms and the closest oxygen atoms of water molecules was maintained. The resulting box, with a volume of 79 nm^3, contained 2371 water molecules and a total of 7681 atoms. This molecular system was used as starting point for two separate MD simulations, referred to in the following with the symbols **A** and **B.**

The equilibration of the BPTI/water system was done in four stages. In order to maintain the original, energy-refined distance geometry NMR structure of BPTI, we applied a harmonic atom position-restraining potential energy term to the BPTI atoms, using the starting structure as reference structure. During 93 steps of energy minimization with the steepest descent method of the full periodic box, the force constant of the position-restraining potential energy term was 9000 kJ mol^{-1} nm^{-2}. Next, three stages of equilibration used the molecular dynamics method with decreasing position-restraining force constant and with increasing temperature. The initial atomic velocities were taken from a Maxwell–Boltzmann distribution at 100 K. In all simulations, the temperatures of the protein and the solvent were separately coupled to a temperature bath.[31] The pressure was kept constant by coupling to a pressure bath[31] at 1.0 bar. Bonds were kept rigid using the SHAKE method[32] with a relative tolerance of 10^{-4}. A cut-off radius of 0.8 nm for the nonbonded interactions was applied. The Coulomb interactions up to 1.2 nm were only calculated every 10 time steps of 0.002 ps each, when the pair list was updated. During the first 2 ps at 100 K, the constants for coupling to the temperature and pressure baths were 0.01 and 0.05 ps, respectively, and the position-restraining force constant was reduced to 900 kJ mol^{-1} nm^{-2}. For the next 2 ps, the temperature was increased to 200 K and the position-restraining force constant decreased to 90 kJ mol^{-1} nm^{-2}. In the third and final stage we simulated for 6 ps at 277 K without any position restraining. The constants for coupling to the temperature and pressure baths were now increased to 0.1 and 0.5 ps, respectively. At this point the time origin was reset to 0, and throughout the following 1.4-ns simulation, which is denoted **A,** the parameters of this final equilibration stage were maintained. For analysis, the coordinates were stored every 0.1 ps.

For the simulation **B** with the inclusion of 642 NOE derived atom–atom distance restraints,[7] the final 6-ps equilibration stage of simulation **A** was repeated, but this time using conventional [$\tau_{dr} = 0$ in (1)] atom–atom distance restraining.[33] The potential energy term was quadratic for distance violations smaller than 0.3 nm, as in (1), and linear beyond that value.[34] The force constant was $k_{dr} = 4000$ kJ mol^{-1} nm^{-2}. The geometric definition of the pseudoatoms[35] and hydrogen atoms involved in the distance restraints is given in refs. 6 and 36. Again the time origin was reset at 0 at the end of the 6 ps equilibration stage, and the same parameters were used for the simulation of the first 200 ps. Then, the simulation was continued by applying time averaging [$\tau_{dr} > 0$ in (1)] to the distance restraints[2,37] for the subsequent 300 ps. When changing to time averaging, the instantaneous distances were used as average distances and the potential energy term V_{phys} was quadratic for all violations. The relaxation time τ_{dr} for the exponential memory function weighting the distances of previous steps was set to 10 ps. After a total simulation period of 500 ps, the distance restraints potential energy term (1) was switched off ($k_{dr} = 0$) for the last 300 ps.

We use the crystal structure form II[9] for comparison with simulated structures. The coordinates of conformation A were chosen for Glu-7 and Met-52. For comparisons with the NMR structures, one of the final, energy-minimized conformers selected to represent the solution structure[7] was used.

RESULTS AND DISCUSSION

Time Dependence of Global Molecular Properties

A number of properties can be used to decide whether a MD simulation has properly equilibrated. They include a levelling off and stability of the potential energy, the radius of gyration and the rmsd of all or a subset of the atomic positions relative to a reference structure. Figure 1 shows the total potential energy and the Lennard–Jones energy for the protein-solvent interaction of simulation **A** (no dis-

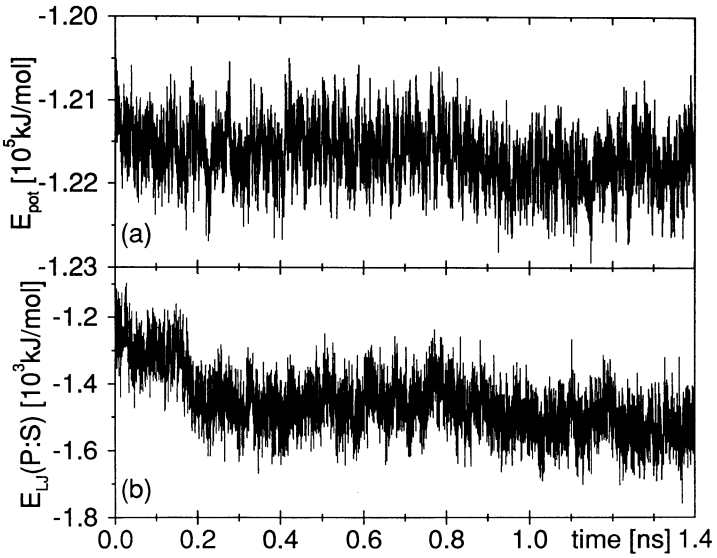

Fig. 1. Total potential energy **(a)** and protein–solvent Lennard–Jones energy **(b)** (in kJ mol^{-1}) as a function of time (in ns) of simulation **A** (without application of distance restraining).

tance restraining). The total potential energy (Fig. 1a) is rather constant. Components of the total energy may show a different behaviour as a function of time. Figure 1b shows that the Lennard–Jones interaction energy between the protein and the solvent exhibits a significant drop around 150–170 ps, from −1300 to −1500 kJ mol^{-1}. In simulation **B** this energy remains quite stable for the first 450 ps. There is a similar drop of this energy between 400 and 550 ps, of approximately 250 kJ mol^{-1}, during the application of time-averaged distance restraints and after the omission of distance restraints (not shown). The Coulomb energies for protein-protein interactions (Fig. 2a) and protein–solvent interactions (Fig. 2b) of simulation **B** not only show much larger rms fluctuations than those for the Lennard–Jones interactions (not shown for simulation **B**), but there is also a pronounced oscillatory behaviour characterized by different time scales: fast energy exchange within 10 ps superimposed on slower exchange over about 150 ps.

For simulation **A** the radius of gyration, determined from 10-ps averages of the trajectory, oscillates with a period of about 100 ps about its mean value of r_G = 1.15 nm (Fig. 3). Throughout simulation **B** the radius of gyration steadily increases but remains within the range found for simulation **A.**

The rms displacements of the C^α atoms and of all atoms calculated as 10-ps averages relative to the starting structure are shown for both simulations in Figure 4a and relative to the crystal structure II[9] in Figure 4b. For simulation A the displacement of atoms relative to the starting structures shows, apart from the usual initial deviation, sudden increase at $t\approx$150 and $t\approx$400 ps for all atoms and at $t\approx$170 ps for the C^α atoms. The graphs obtained from simulation **B** display significant steps at $t\approx$300 psec for all atoms and at $t\approx$500 ps for the C^α atoms. The latter coincides with the removal of the distance restraints. In simulation **A** the protein can move freely within the accessible configurational space defined by the force field, whereas this space is partially restricted by inclusion of the distance-restraining force field term in simulation **B.** This is illustrated by the significantly smaller rms displacements during the first 500 ps of simulation **B,** when distance restraints are still applied. The displacement of all atoms relative to the starting structure is about 25% smaller and for the C^α atoms it is about halved in simulation **B.** Relative to the crystal structure (Fig. 4b), this effect is less pronounced. In all cases except for the all-atom displacement relative to the starting structure, the curves converge to similar values for simulations **A** and **B.** Moreover, the average displacement values from the comparison to two different (NMR, X-ray) reference structures differ only by about 0.05 nm. This means that in the simulations one or more conformational regions are accessed that are approximately equally "far away" from single structure representations derived from X-ray diffraction and from NMR spectroscopy. This clearly indicates molecular flexibility and requires a more detailed analysis of the conformational regions accessible to the protein. Figures 1a, 2, and 4 show that the potential energy, the radius of gyration and the r.m.s. atomic positional deviation from two reference (NMR, X-ray) structures have more or less equilibrated after approximately 200–400 ps. We note that such an equilibration period is long compared with the complete simulation period of MD simulations of proteins in solution published so far.[22,23,38–40]

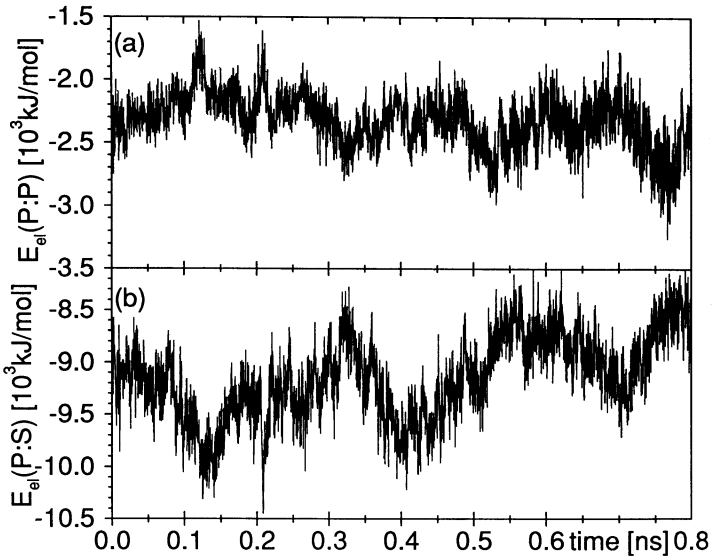

Fig. 2. Coulomb energies (in kJ mol^{-1}) for the protein–protein **(a)** and the protein–solvent **(b)** interactions as function of time (in ns) for simulation **B** with instantaneous (0–200 ps), time-averaged (200–500 ps), and omission of (500–800 ps) distance restraints.

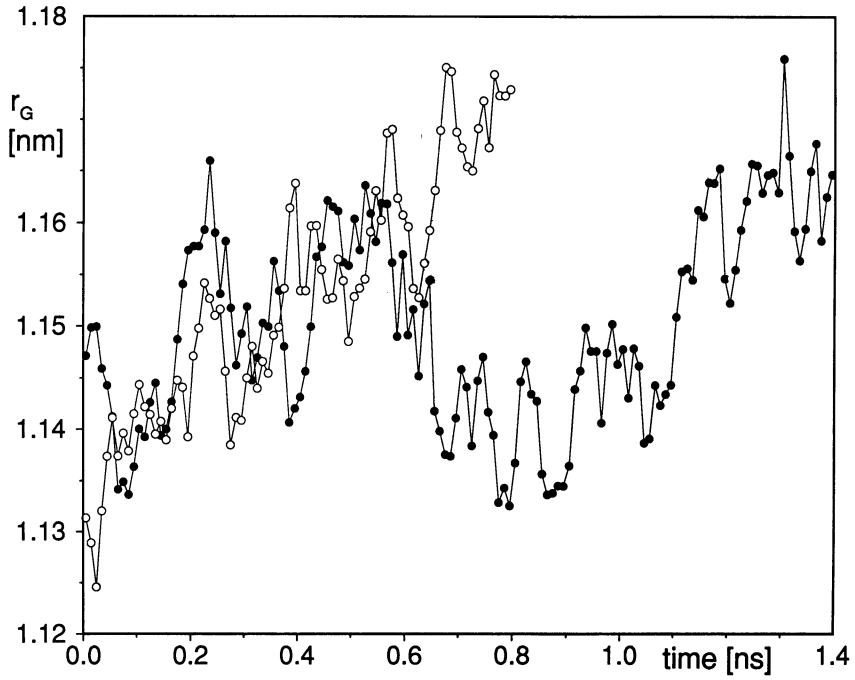

Fig. 3. Variation of 10-psec averages of the radius of gyration r_G (in nm) of BPTI with time (in ns) for simulation **A** (●) and simulation **B** (○) with instantaneous (0–200 ps), time-averaged (200–500 ps), and omission of (500–800 ps) distance restraints.

Characterization of Conformations

In order to characterize changes occurring within the BPTI structure, we calculated averages of the ϕ and ψ dihedral angles over different parts of the trajectories, which were characterized by different ϕ, ψ angle values. The trajectories show many ϕ, ψ dihedral angle changes. However, only when a dihedral angle resides in a specified potential energy minimum for a certain minimal time span would we like to consider this as a distinct conformation. The longer-lived conformations can be detected as follows. Whenever two distinct dihedral angle conformers both contribute substantially to a trajectory average, the rms dihedral angle fluctuation will be

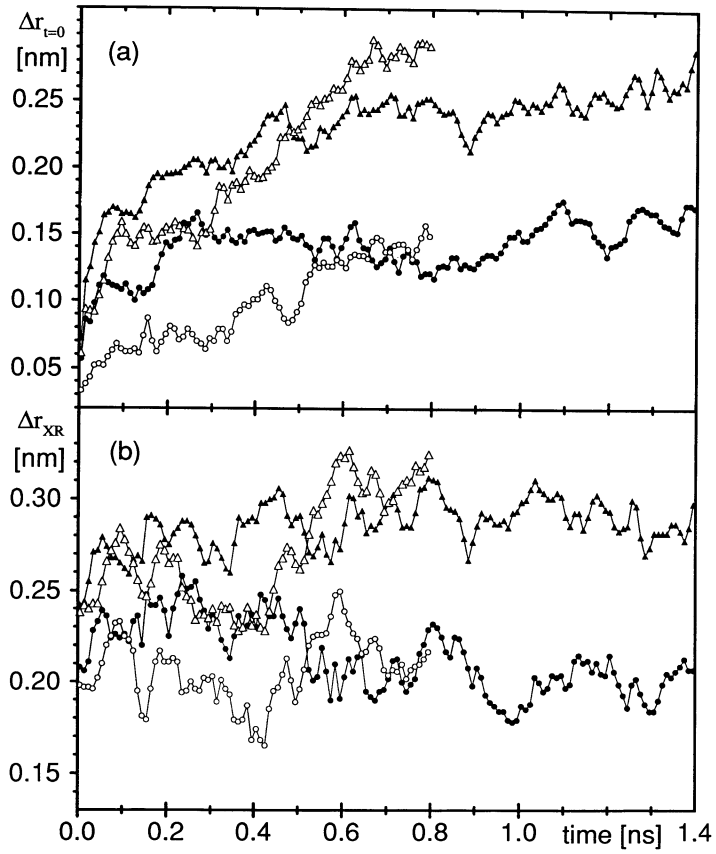

Fig. 4. Variation with time of the rms displacement of atom positions (in nm) averaged over 10-ps subtrajectories relative to **(a)** the starting structure and **(b)** the crystal structure II. The filled symbols represent the results obtained from simulation **A**, the open symbols those from simulation **B** with instantaneous (0–200 ps), time-averaged (200–500 ps), and omission of (500–800 ps) distance restraints. Circles denote the mean atomic displacement for C^{α} atoms (also used for the least-squares fit of structures), and triangles for all atoms.

large. We took fluctuations larger than 30° as indicative for the occurrence of backbone angle transitions and analyzed them further. We found that only ψ and ϕ angles of consecutive residues flip together at a given time (i.e., a rotation of the plane of the peptide bond takes place). The residue pairs Leu-6-ψ Glu-7-ϕ, and Arg-17-ψ Ile-18-ϕ (Fig. 5a) undergo such transitions in both simulations **A** and **B.** In addition, we find transitions in simulation **A** for Gly-37-ψ Cys-38-ϕ and in simulation **B** for Pro-13-ψ Cys-14-ϕ (Fig. 5b). Some other backbone angles exhibit a slow drift of 30–50° over 500 ps, or show large fluctuations around a mean value with short life-times of just a few picoseconds in either extreme. These are not considered to be characteristic for defining conformations but will be accounted for in the description of mean angle values and rms fluctuations. Using the six backbone angles of each simulation which lead to conformational transitions as defined here, we define four periods within the trajectory of simulation **A** and three periods within simulation **B** with small variations in backbone angles for which an average conformation can be defined. Conformation **A1** is averaged over 50–100 ps of simulation **A, A2** over 400–800 ps, **A3** over 900–1050 ps, and **A4** over 1100–1400 ps, respectively. Similarly, conformations **B1, B2,** and **B3** are averaged over 100–200, 300–500, and 610–750 ps of simulation **B.**

The average ϕ and ψ angles of all conformations and of the X-ray and the NMR structure are displayed in Figure 6 and the rms fluctuations about the averages in Figure 7. Apart from those pairs of residues mentioned above and the N- and C-terminal residues the dihedral backbone angles differ little among the 9 structures. The largest deviations of dihedral angles are found between the X-ray and the NMR conformer selected for the present comparisons, but, interestingly, just for angles that are either identical with or next to the dihedrals that distinguish conformations in either of the simulations (Table I): the dihedral angles of Pro-13-ψ, Ala-16-ϕ, Gly-36-ψ, Gly-37-ϕ, and Cys-55-ψ differ by 60° or more.

It is not surprising to find most of the transitions and the largest structural differences for residues of

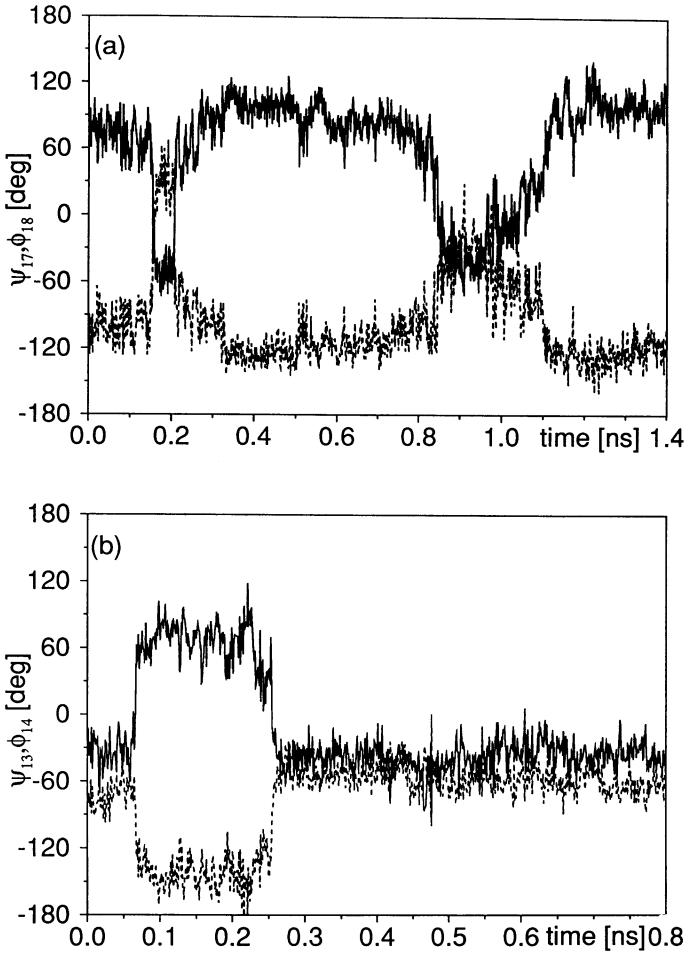

Fig. 5. Variation of backbone dihedral angles (in degrees) with time (in ns). The angles Arg-17-ψ (solid) and Ile-18-ϕ (dotted) of simulation **A** are given in **(a)**, the angles Pro-13-ψ (solid) and Cys-14-ϕ (dotted) of simulation **B** in **(b)**.

the "flexible top" of BPTI, spanning residues 12–16 and 36–38. However, several transitions involve residues that are at the ends of regions with well defined secondary structure, i.e., the 3_{10}-helix spanning residues 3–6, the antiparallel β-sheet involving residues 18–24 and 29–35, and an α-helix of residues 47–56. Residues at the ends of secondary structure elements sometimes show even larger flexibility than those which do not participate in secondary structures at all. Figures 7a and b display the rms fluctuations for the backbone dihedral angles of conformations from simulation **A.** Besides the chain termini there are four regions with a larger dihedral angle mobility: 3–6, 14–18, 24–30, and 36–38. For the conformations from simulation **B** (Fig. 7c and d) large fluctuations are observed only for the two regions 6–7 and 36–38. These differences between dihedral angle fluctuations from simulation **A** and from simulation **B** are partly reflected in the plots for the rms positional fluctuations of the α- and γ-carbon atoms (Figs. 8 and 9). Local maxima in the plots for simulation **A** can also be found for simulation **B,** but the size of the fluctuations is generally smaller in the partially restrained simulation **B.** There exists only qualitative agreement between the positional fluctuations obtained from the simulations and those calculated from the crystallographic *B*-factors (bold line) according to Eq. (2):

$$\langle(\Delta r)^2\rangle^{1/2} = [3B/(8\pi^2)]^{1/2} \qquad (2)$$

The agreement with the crystallographic *B*-factors is slightly better for the sidechain atoms than for the backbone atoms. This may be explained by the shorter time scales of the side chain motions which allow for better sampling within the averaging period. We note that averaging periods of a few hundred picoseconds are too short to sample accurately the relatively rare transitions of backbone angles.

Table II summarizes various properties calculated from the coordinates of crystal structure form II, XR,

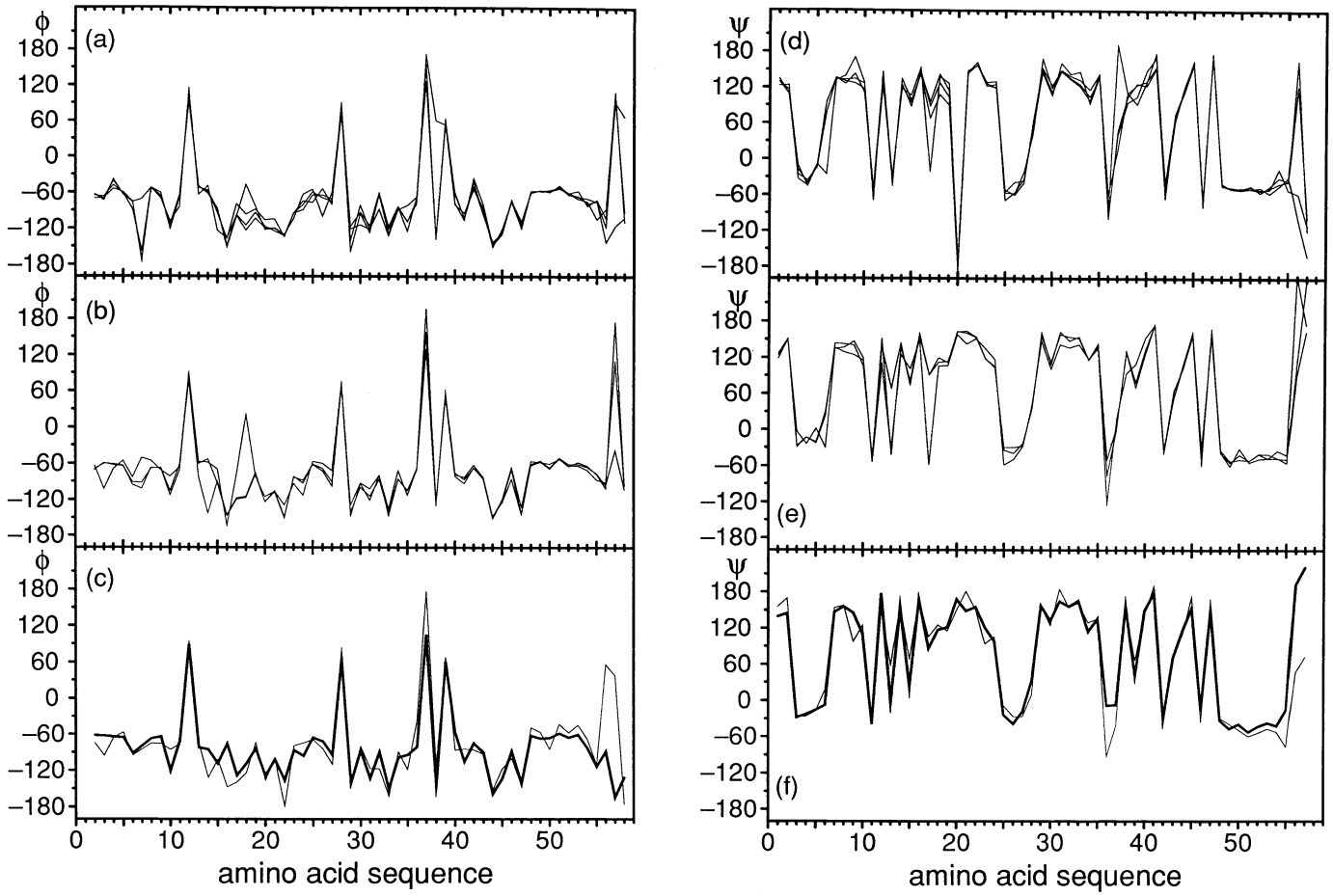

Fig. 6. Backbone dihedral angle values (in degrees) for the various conformations from the simulations **A** and **B**. The ϕ angles averaged over the subtrajectories **A1–A4** from simulations **A** are given in **(a)**, those averaged over **B1–B3** of simulation **B** in **(b)**, and the ϕ angles of the crystal structure II (bold) and the NMR solution structure in **(c)**. The corresponding ψ angles are represented in **(d)**, **(e)**, and **(f)**.

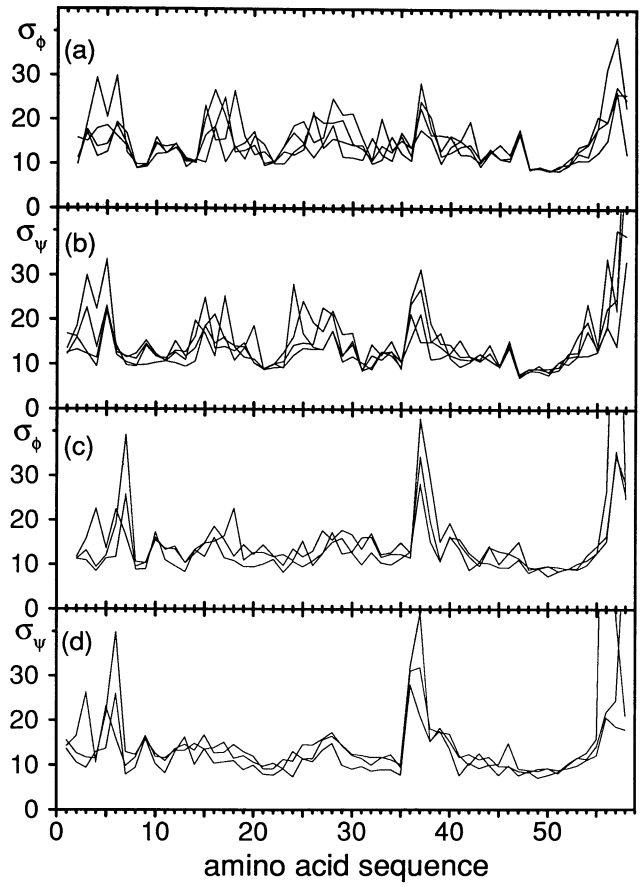

Fig. 7. The root mean square fluctuations of the backbone dihedral angles (in degrees) for the various conformations from the simulations **A** and **B**. The fluctuation of the ϕ angles in **(a)** and the ψ angles in **(b)** are averaged over the subtrajectories **A1–A4** of simulation A; the fluctuations of the ϕ and ψ angles over the subtrajectories **B1–B3** of simulation **B** are given in **(c)** and **(d)**.

the distance geometry refined NMR structure, NMR, and the conformations **A1–A4** and **B1–B3** obtained from positional averaging over the corresponding parts of the trajectory. The solvent accessible surface area F, the number of hydrogen bonds involved in the anti-parallel β-sheet, $N_{\text{H-bonds}(\beta)}$, in the α-helix, $N_{\text{H-bonds}(i,i+4)}$, and in the 3_{10}-helix, $N_{\text{H-bonds}(i,i+3)}$, and the assignment of residues to any of the secondary structures were calculated from the subtrajectory-averaged coordinates using the analysis program DSSP.[41] The radius of gyration r_G, dihedral angle fluctuations, and distance restraint information were calculated directly from the subtrajectories.

The significantly smaller dihedral angle and positional fluctuations for **A1** and **B1** partially arise from averaging over relatively short subtrajectories and the immediate penalization of distance-bound violations. However, the mobility within the subtrajectory **B2,** while time-averaging of distance restraints is applied, is of the same order as that found within simulation **A** (without distance restraining). Yet, the average distance restraint violation ⟨dr-viol⟩, the total number of distance violations N_{viol}, and the number of distance violations larger than 0.1 nm, $N_{\text{viol}}(>0.1)$, are almost equal for conformations **B1, B2,** and the NMR structure. This shows that the use of time averaging of distance restraints allows for an appropriate description of protein mobility[2] while fulfilling atom–atom distance bounds. Since the radius of gyration, r_G, is quite variable (Fig. 3), the solvent accessible surface area, F, may be a better quantity to judge the compactness and stability of the tertiary fold as well as the correlation with NOE violations. The surface area of conformations with many and large NOE violations is about 2–3 nm^2 larger than that of conformations with fewer and smaller NOE violations. The regions with defined secondary structure are relatively stable as can be seen from the number of hydrogen bonds involved therein and from the range of residues in these structures. However, the ends of the 3_{10}-helix and the β-strands are somewhat disordered, as was already discussed above in the context of the backbone angle transitions.

TABLE I. Dihedral Angle Values*

Angle	XR	NMR	A1	A2	A3	A4	B1	B2	B3
Leu-6-ψ	−8	+17	−26	+79	+95	+96	+22	+29	−31
Glu-7-ϕ	−79	−85	−72	−158	−172	−176	−92	−100	−50
Pro-13-ψ	−3	+59	−42	−36	−32	−32	+69	−42	−34
Cys-14-ϕ	−85	−132	−49	−57	−63	−59	−143	−53	−58
Lys-15-ψ	+22	+69	+98	+86	+106	+95	+74	+81	+100
Ala-16-ϕ	−76	−148	−138	−137	−153	−149	−146	−148	−164
Arg-17-ψ	+86	+106	+68	+87	−21	+96	+90	+92	−59
Ile-18-ϕ	−109	−124	−97	−115	−48	−124	−115	−118	+21
Gly-36-ψ	−8	−93	−58	−96	−100	−86	−49	−74	−125
Gly-37-ϕ	+105	+177	+127	+156	+154	+172	+129	+157	−162
Gly-37-ψ	−7	−42	+15	+39	+46	−168	+4	−1	+17
Cys-38-ϕ	−146	−110	−137	−138	−135	+60	−130	−131	−123
Cys-55-ψ	−15	−77	−37	−37	−31	−52	−55	−50	−42

*Dihedral angle values (in degrees) for the crystal structure II (XR), the energy-refined distance geometry NMR conformer selected for the present comparisons (NMR), and the conformations averaged over the trajectory parts of the MD simulation **A** without any distance restraining (**A1–A4**) and of the MD simulation **B** with instantaneous (B1), then time-averaged (**B2**), and finally without (**B3**) distance restraining. The ψ, ϕ dihedral angle pairs that show transitions of more than 60° in the simulations or show a deviation of more than 60° from the XR or NMR conformers are listed.

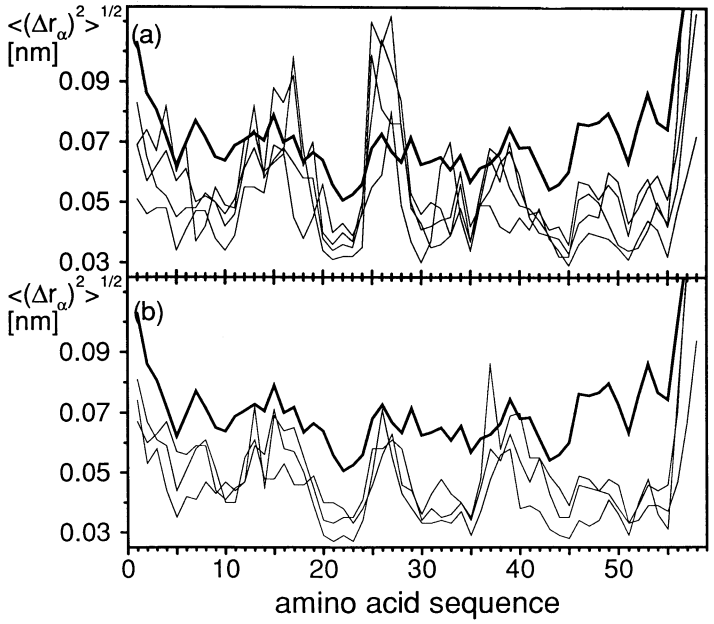

Fig. 8. Positional root mean square fluctuations (in nm) of the C^{α} atoms as calculated from the subtrajectories **A1–A4** of simulation **A (a)** and from **B1–B3** of simulation **B (b).** The positional fluctuations as obtained from the crystallographic *B*-factors of crystal structure II using Eq. (2) are included for comparison (bold line).

Comparison of the Backbone Conformations in the Crystalline State, in Solution and From Molecular Dynamics Simulations

We compared the 9 selected structures with each other on the basis of dihedral angle differences (Table III) for residues 3–56 and positional rms deviations (Table IV) for residues 1–56. Any pair of average structures picked from the same simulation has very similar values for these quantities, especially if they are taken from consecutive parts of the same trajectory. Simulations **A** and **B** started from almost identical structures (only the last stage in the equilibration procedure was different) directly derived from NMR data. Therefore, it is not surprising that **A1, B1,** and NMR resemble each other quite well. The structures **B2** and **B1** are closest to both of the experimental structures, XR and NMR. The good agreement with NMR could be expected since the NMR restraints were used in simulation **B.** The good agreement with the crystal structure II by the same structures is surprising. Although the two experimental structures are similar with respect to the

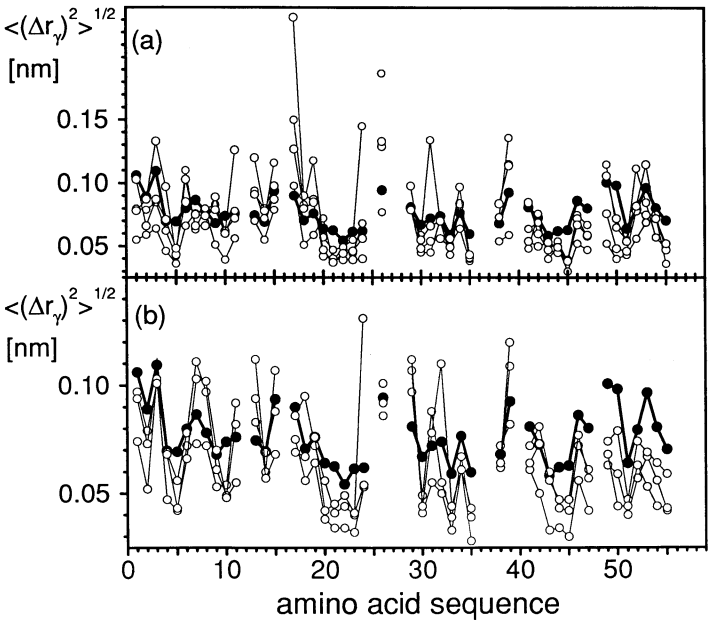

Fig. 9. Positional root mean square fluctuations (in nm) of all γ carbon atoms as calculated from the subtrajectories **A1–A4** of simulation **A (a)** and from **B1–B3** of simulation **B (b).** The positional fluctuations as obtained from the crystallographic *B*-factors of crystal structure II using Eq. (2) are included for comparison (●).

positional rms deviation values, they certainly are not indistinguishable from each other, and they have larger average backbone angle differences and positional rms deviations between each other than to either of the simulation structures **B1** and **B2.**

In general, we find that the structure of BPTI is maintained throughout simulations **A** and **B.** Even if additional distance information is not included in the simulations, the fit of the simulated structures to the solution and the crystal structures remains almost as good as the fit between the experimental structures themselves. Regions of the protein that are intrinsically more flexible, such as side chains on the protein surface and the N- and C-termini, will tend to explore a greater region of conformation space if they are not restrained by additional information. The increased mobility in these regions does not destroy the secondary or tertiary structure but accounts for the possibility of conformational variations that can cause differences between different crystal structures and solution structures.

Comparison of Conformational Mobility in the Solid State, in Solution and From Molecular Dynamics Simulations

Thorough comparisons of the structures of the three available crystal forms, and between these crystal structures and the NMR solution structure have been presented previously.[7,9–11] The overlap of the backbone atoms of the crystal structures is very good. Only the backbone atoms around Lys-15 and Lys-26 show deviations larger than 0.1 nm. In crystal structure II[9] the sidechains of Glu-7 and Met-52 were found to have two different conformations each, and in crystal structure III[10] two conformations were found for Arg-39 and Asp-50. The overall agreement of the average NMR structure with the three crystal structures is very good, with a rms deviation[7] of the backbone atoms of 0.124 nm for residues 2–56. Significant deviations could be located for residues 25–26 and 46–48.

The rms fluctuations of the backbone dihedral angles of Lys-15, Ala-25, and Lys-26 for the conformations from simulation **A** (Fig. 7a and b) are larger than for most of the other regions, but transitions into different conformations were neither found for these angles nor for those of residues 46–48. Also, the positional rms fluctuations of the C^{α} atoms (Fig. 8) are much increased in these regions. Therefore, the variations with the sets of crystal and solution structures may be explained purely on the basis that larger atom movements are possible for residues in these regions.

The γ-carbon atoms of the residues with multiple conformations in the crystal structures, Glu-7, Arg-39, and Met-52, exhibit large positional fluctuations (Fig. 9). The side chain dihedral angles of Arg-39 and Met-52 undergo many transitions in simulation **A,** and side chain conformations which resemble the crystal conformations are observed. The sidechain dihedral angles of Glu-7 are very stable throughout simulation **A,** and χ^1 quickly adopts a value of ap-

TABLE II. Comparison of Properties*

	XR	NMR	A1	A2	A3	A4	B1	B2	B3
t_{start} (ps)			50	400	900	1100	100	300	610
Length (ps)			50	400	150	300	100	200	140
$\sigma(\phi)$ (°)			14	17	15	16	14	20	16
$\sigma(\psi)$ (°)			13	19	16	17	15	23	15
$\sigma(\chi)$ (°)			22	35	26	38	17	35	24
σ(all) (°)			15	25	21	24	15	27	19
$\langle(\Delta r_\alpha)^2\rangle^{1/2}$ (nm)			.047	.063	.058	.061	.049	.058	.052
$\langle(\Delta r_\gamma)^2\rangle^{1/2}$ (nm)			.059	.086	.081	.078	.061	.076	.069
⟨dr-viol⟩ (nm)	.008	.001	.013	.020	.029	.035	.001	.001	.016
N_{viol}	88	72	141	165	186	193	54	66	155
$N_{viol}(>0.1)$	13	0	29	45	74	75	0	0	30
r_G (nm)	1.12	1.12	1.14	1.15	1.15	1.16	1.14	1.15	1.17
F (nm^2)	39.8	41.2	42.0	43.5	45.1	45.8	42.5	41.9	44.3
$N_{H\text{-}bonds(\beta)}$	10	10	10	9	8	10	10	10	9
$N_{H\text{-}bonds(i,i+4)}$	7	6	7	7	7	7	7	7	7
$N_{H\text{-}bonds(i,i+3)}$	3	4	2	2	2	2	3	3	2
3_{10}-Helix (res)	3–6	4–6	4–6	3–5	3–5	3–5	3–6	3–6	4–6
α-Helix (res)	48–55	48–54	48–55	48–55	48–55	48–55	48–55	48–55	48–55
β-Sheet† (res)	18–24	18–24	18–24	18–23	19–23	18–24	18–24	18–24	19–24
(res)	35–29	35–29	35–29	35–30	34–30	35–29	35–29	35–29	34–29

*Comparison of properties of the crystal structure form II (XR), the energy-refined distance geometry NMR conformer selected for the present comparisons (NMR), and conformations from the MD simulation **A** without any distance restraining **(A1–A4)** and of MD simulation **B** with instantaneous **(B1)**, then time-averaged **(B2)**, and finally without **(B3)** distance restraining. σ, rms fluctuation; $\langle(\Delta r)^2\rangle^{1/2}$, atomic positional fluctuation; ⟨dr-viol⟩, average distance violation; N_{viol}, total number of distance violations; $N_{viol}(>0.1)$, number of distance violations larger than 0.1 nm; r_G, radius of gyration; F, surface area; $N_{H\text{-}bonds}$, number of hydrogen bonds observed in the β-sheet, the α-, and the 3_{10}-helix.
†For all conformations there is also an antiparallel β-strand contact from residue 45 to residue 21

TABLE III. Backbone Dihedral Angle Differences*

	XR	NMR	A1	A2	A3	A4	B1	B2	B3
XR	—	18.5	17.0	19.2	22.2	22.9	12.5	12.5	12.8
NMR	5	—	24.3	23.5	27.1	26.4	14.6	15.9	25.3
A1	2	5	—	12.1	14.9	17.7	15.4	13.0	15.9
A2	8	8	2	—	11.1	11.9	14.8	11.7	15.2
A3	9	9	4	3	—	11.9	18.9	16.5	14.9
A4	11	8	5	3	4	—	20.0	17.2	19.7
B1	3	5	2	3	7	7	—	6.9	18.1
B2	3	3	0	0	5	5	2	—	15.9
B3	10	9	5	5	3	6	7	4	—

*Backbone dihedral angle differences between the crystal structure II (XR), the energy-refined distance geometry NMR conformer selected for the present comparisons (NMR), and average conformations from the MD simulation **A** without any distance restraining **(A1–A4)** and from the MD simulation **B** with instantaneous **(B1)**, time-averaged **(B2)**, and finally without **(B3)** distance restraining. The upper right triangle gives the average dihedral angle deviations for the ϕ,ψ angles (in degrees) of residues 3 to 56, and the lower left triangle the number of these ϕ,ψ dihedral angles differing by more than 60°.

proximately −170°, which differs from any of the angles in the crystal structures by more than 100°. Also the χ^1 angle of Asp-50 is stable around 180° and quite different from any of the crystal structure conformations.

CONCLUSIONS

The long-time simulations of BPTI in solution revealed an often oscillatory variation of properties as global as the protein–protein or protein–solvent Lennard–Jones and Coulomb interaction energies. For the radius of gyration and the average rms displacement of either backbone atoms or of all atoms, oscillations are observed with time scales ranging from tens to hundreds of picoseconds. Due to poor statistics, it is difficult to judge if the slower oscillations (> 200 ps) resemble a realistic behavior or represent only the relaxation of very slowly equilibrating motions. To answer this question even longer simulations need to be performed.

It took up to 300 ps for some quantities to reach a fairly stable equilibrium value (neglecting oscillations), which is much longer than what is considered to be sufficient for equilibration in previously reported MD simulations. Especially the time behavior of the Coulomb energies ought to give reason to reconsider the validity of a free energy calculation based on short time simulations for large systems.[25]

TABLE IV. Root Mean Square Deviations*

	XR	NMR	A1	A2	A3	A4	B1	B2	B3
XR	—	.088	.117	.123	.126	.141	.074	.077	.127
NMR	.157	—	.133	.145	.148	.162	.083	.089	.156
A1	.178	.221	—	.069	.091	.114	.087	.089	.099
A2	.189	.229	.141	—	.079	.077	.104	.105	.092
A3	.211	.235	.178	.131	—	.068	.114	.111	.120
A4	.214	.237	.202	.135	.121	—	.132	.131	.124
B1	.150	.162	.146	.163	.174	.198	—	.038	.102
B2	.141	.171	.149	.144	.168	.188	.088	—	.097
B3	.211	.261	.191	.146	.194	.201	.184	.152	—

*Rms deviations (in nm) between the crystal structure form II (XR), the energy-refined distance geometry NMR conformer selected for the present comparisons (NMR), and average conformations from the MD simulation **A** without any distance restraining (**A1–A4**) and from the MD simulation **B** with instantaneous (**B1**), time-averaged (**B2**), and finally without (**B3**) distance restraining. The upper right triangle gives the deviations between C^{α} atoms and the lower left triangle between all atoms for residues 1–56.

With respect to additional structural information we find that the dynamic behavior as derived from dihedral angle fluctuations and positional rms fluctuations can also be reproduced reliably from short time simulations. But many backbone and side chain dihedral angles undergo transitions with a rate of one per hundreds of picoseconds. In order to correlate these transitions with the structural and dynamic information obtainable by NMR spectroscopy with its much longer time scale, the simulation length of 1.4 ns is still much too short to provide significant statistical information. However, the backbone angle transitions occur in regions that are either characterized by or indicative of increased flexibility. They provide a clear indication that larger flexibility must not only be accounted for by somewhat larger variation of a dihedral angle or of an atom position around their mean values, but also by real dihedral angle transitions in which actual potential energy barriers are crossed. Even a protein as stable as BPTI may undergo these conformational backbone transitions. However, it is also seen from this study that backbone angle transitions do not occur randomly, but within intrinsically flexible regions and in a rather concerted manner. The simultaneous transition of two consecutive ψ and ϕ angles ensures that the disturbance of the overall backbone fold is very minor. Of special interest is the relatively poor conformational stability of residues at the ends of regular secondary structure regions when compared to other regions that have no defined secondary structure at all. Here we find, that the participation of residues at either end of the secondary structure regions is a dynamic process and does not affect the overall fold of the backbone.

The molecular model, atomic interaction function, and computational procedure that were used generate BPTI trajectories which are in agreement with the available experimental data. The rms deviation of the structure averaged over the subtrajectory from 1100 to 1400 ps relative to the crystal structure II is 0.14 nm for the C^{α} atoms and 0.21 nm for all atoms, which is only slightly larger than the corresponding rms deviations of 0.09 and 0.16 nm, between the NMR structure and the crystal structure II, i.e., 0.05 nm in terms of C^{α} rmsd values, or 3° in terms of backbone dihedral angles. The average NOE distance violations are only 0.03 nm. The secondary structure elements are maintained in the simulations. The long simulations did not give indications of deficiencies in the GROMOS force field.

By the application of atom–atom distance restraints derived from NOE data as a distance-restraining term in the potential energy function, the quality of the obtained trajectory could be improved: the simulated BPTI structures now are closer to both the NMR solution structure and the X-ray crystal structure than the difference between these experimental structures. The application of time-averaged distance restraints with an averaging time of τ_{dr} = 10 ps yields a slightly better representation of the experimental information than either the NMR solution structure or the trajectory generated without time averaging: it yields less large (> 0.1 nm) NOE violations and its structural properties are closer to those of the well-defined X-ray crystal structure, while the atomic mobility as observed in nonrestrained MD simulations is maintained.

The detailed analysis of the BPTI motions at the atomic level reveals the largest flexibility of BPTI around those residues (12–16, 36–38) for which the largest structural difference between the NMR solution structure and the X-ray crystal structure is observed. A more surprising observation is that the residues at the ends of secondary structure elements show sometimes a larger mobility than those in loops of the protein.

The presently described simulations of BPTI show that sufficient agreement with the experimental data can be obtained to allow for an interpretation of protein dynamics on the atomic level within a picosecond to nanosecond time scale.

ACKNOWLEDGMENTS

The simulations were partly performed on the NEC SX-3/22 of the Centro Svizzero di Calcolo Sci-

entifico in Manno, Ticino, Switzerland. We thank G. Otting for helpful discussions. This work was supported by the Schweizerischer Nationalfonds (project 31.32033.91).

REFERENCES

1. Wüthrich, K. "NMR of Proteins and Nucleic Acids." New York: Wiley, 1986.
2. Torda, A. E., Scheek, R. M., van Gunsteren, W. F. Time-averaged nuclear Overhauser effect distance restraints applied to Tendamistat. J. Mol. Biol. 214:223–235, 1990.
3. Tropp, J. Dipolar relaxation and nuclear Overhauser effects in nonrigid molecules: The effect of fluctuating internuclear distances. J. Chem. Phys. 72:6035–6043, 1980.
4. van Gunsteren, W. F., Berendsen, H. J. C. Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry. Angew. Chem. Int. Ed. Engl. 29:992–1023, 1990.
5. van Gunsteren, W. F., Mark, A. E. On the interpretation of biochemical data by molecular dynamics computer simulation. Eur. J. Biochem. 204:947–961, 1992.
6. van Gunsteren, W. F., Berendsen, H. J. C. Groningen Molecular Simulation (GROMOS) Library Manual, Biomos, Groningen, 1987.
7. Berndt, K. D., Güntert, P., Orbons, L. P. M., Wüthrich, K. Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. J. Mol. Biol. 227:757–775, 1992.
8. Deisenhofer, J., Steigemann, W. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1·5 Å resolution. Acta Crystallogr. Sect. B 31:238–250, 1975.
9. Wlodawer, A., Walter, J., Huber, R., Sjölin, L. Structure of bovine pancreatic trypsin inhibitor: Results of joint neutron and X-ray refinement of crystal form II. J. Mol. Biol. 180:301–329, 1984.
10. Wlodawer, A., Nachman, J., Gilliland, G. L., Gallagher, W., Woodward, C. Structure of form III crystals of bovine pancreatic trypsin inhibitor. J. Mol. Biol. 198:469–480, 1987.
11. Wlodawer, A., Deisenhofer, J., Huber, R. Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. J. Mol. Biol. 193:145–156, 1987.
12. Otting, G., Liepinsh, E., Wüthrich, K. Disulfide bond isomerization in BPTI and BPTI(G36S): An NMR study of correlated mobility in proteins. Biochemistry 32:3571–3582, 1993.
13. Richarz, R., Nagayama, K., Wüthrich, K. Carbon-13 nuclear magnetic resonance relaxation studies of internal mobility of the polypeptide chain in basic pancreatic trypsin inhibitor and a selectively reduced analogue. Biochemistry 19:5189–5196, 1980.
14. Tüchsen, E., Woodward, C. Hydrogen exchange kinetics of peptide amide protons at the bovine pancreatic trypsin inhibitor protein-solvent interface. J. Mol. Biol. 185:405–419, 1985.
15. Wagner, G., Brühwiler, D. Toward the complete assignment of the carbon nuclear magnetic resonance spectrum of the basic pancreatic trypsin inhibitor. Biochemistry 25: 5839–5843, 1986.
16. Wagner, G., Wüthrich, K. Dynamic model of globular protein conformations based on NMR studies in solution. Nature (London) 275:247–248, 1978.
17. Wagner, G., Wüthrich, K. Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra: Basic pancreatic trypsin inhibitor. J. Mol. Biol. 155:347–366, 1982.
18. McCammon, J. A., Gelin, B. R., Karplus, M. Dynamics of folded proteins. Nature (London) 267:585–590, 1977.
19. van Gunsteren, W. F., Karplus, M. Protein dynamics in solution and in a crystalline environment: A molecular dynamics study. Biochemistry 21:2259–2274, 1982.
20. van Gunsteren, W. F., Berendsen, H. J. C. Computer simulation as a tool for tracing the conformational differences between proteins in solution and in the crystalline state. J. Mol. Biol. 176:559–564, 1984.
21. van Gunsteren, W. F., Berendsen, H. J. C., Hermans, J., Hol, W. G. J., Postma, J. P. M. Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray data. Proc. Natl. Acad. Sci. U.S.A. 80:4315–4319, 1983.
22. Levitt, M., Sharon, R. Accurate simulation of protein dynamics in solution. Proc. Natl. Acad. Sci. U.S.A. 85:7557–7561, 1988.
23. Daggett, V., Levitt, M. A model of the molten globule state from molecular dynamics simulations. Proc. Natl. Acad. Sci. U.S.A. 89:5142–5146, 1992.
24. Brunne, R. M., Liepinsh, E., Otting, G., Wüthrich, K., van Gunsteren, W. F. Hydration of proteins: A comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. J. Mol. Biol. 231:1040–1048, 1993.
25. Smith, P. E., Brunne, R. M., Mark, A. E., van Gunsteren, W. F. Dielectric properties of trypsin inhibitor and lysozyme calculated from molecular dynamics simulations. J. Phys. Chem. 97:2009–2014, 1993.
26. Smith, P. E., van Gunsteren, W. F. Translational and rotational diffusion of proteins. J. Mol. Biol. 236:629–636, 1994.
27. Otting, G., Liepinsh, E., Wüthrich, K. Protein hydration in aqueous solution. Science 254:974–980, 1991.
28. Otting, G., Liepinsh, E., Wüthrich, K. Proton exchange with internal water molecules in the protein BPTI in aqueous solution. J. Am. Chem. Soc. 113:4363–4364, 1991.
29. Richarz, R., Wüthrich, K. High-field ^{13}C nuclear magnetic resonance studies at 90.5 MHz of the basic pancreatic trypsin inhibitor. Biochemistry 17:2263–2269, 1978.
30. Berendsen, H. J. C., Grigera, J. R., Straatsma, T. P. The missing term in effective pair potentials. J. Phys. Chem. 91:6269–6271, 1987.
31. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., Haak, J. R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684–3690, 1984.
32. Ryckaert, J.-P., Ciccotti, G., Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J. Comput. Phys. 23:327–341, 1977.
33. van Gunsteren, W. F., Kaptein, R., Zuiderweg, E. R. P. Use of molecular dynamics computer simulations when determining protein structure by 2D NMR. In: "Proceedings NATO/CECAM Workshop on Nucleic Acid Conformation and Dynamics." Olsen, W. K. (ed.). Orsay, France: CECAM, 1984: 79–92.
34. Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., van Gunsteren, W. F. A protein structure from nuclear magnetic resonance data: lac repressor headpiece. J. Mol. Biol. 182:179–182, 1985.
35. Wüthrich, K., Billeter, M., Braun, W. Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. J. Mol. Biol. 169:949–961, 1983.
36. van Gunsteren, W. F., Boelens, R., Kaptein, R., Scheek, R. M., Zuiderweg, E. R. P. An improved restrained molecular dynamics technique to obtain protein tertiary structure from nuclear magnetic resonance data. In: "Molecular Dynamics and Protein Structure." Hermans, J. (ed.). Western Springs: Polycrystal Book Service, 1985: 92–99.
37. Torda, A. E., Scheek, R. M., van Gunsteren, W. F. Time-dependent distance restraints in molecular dynamics simulations. Chem. Phys. Lett. 157:289–294, 1989.
38. Ahlström, P., Teleman, O., Kördel, J., Forsén, S., Jönsson, B. A molecular dynamics simulation of bovine calbindin D_{9k}: Molecular structure and dynamics. Biochemistry 28: 3205–3211, 1989.
39. de Vlieg, J., Berendsen, H. J. C., van Gunsteren, W. F. An NMR based molecular dynamics simulation of the interaction of the lac repressor headpiece and its operator in aqueous solution. Proteins 6:104–127, 1989.
40. Harte Jr., W. E., Swaminathan, S., Mansuri, M. M., Martin, J. C., Rosenberg, I. E., Beveridge, D. L. Domain communication in the dynamical structure of HIV-1 protease. Proc. Natl. Acad. Sci. U.S.A. 87:8864–8868, 1990.
41. Kabsch, W., Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637, 1983.