Molecular Dynamics of Proteins with the OPLS Potential Functions. Simulation of the Third Domain of Silver Pheasant Ovomucoid in Water

Julian Tirado-Rives and William L. Jorgensen*

Contribution from the Department of Chemistry, Purdue University, West Lafayette, Indiana 47907. *Received May 15, 1989*

Abstract: A molecular dynamics simulation using the OPLS nonbonded potential functions has been carried out for the third domain of silver pheasant ovomucoid in aqueous solution. Insights have been obtained on the quality of the force field, the convergence of such calculations, differences in the protein's structure in the crystal and in aqueous solution, protein hydration, and the dynamics of water molecules near a protein. The simulation covered 100 ps at 25 °C, which allowed complete equilibration prior to averaging and analysis of the results. Continuous monitoring of the potential energy, root-mean-square deviations from the crystal structure, and other properties indicated that convergence to a stable structure was achieved after 30–40 ps. The RMS deviation of the instantaneous structure from the crystal structure after 100 ps is 1.43 Å for the backbone atoms of residues 8–56 and 1.61 Å for all residues. There is substantial reorganization of hydrogen bonds that do not involve secondary structure in comparing the crystal and solution structures, though in the simulation Ala-44 is displaced from the α-helix and Lys-29, Thr-30, Tyr-31, and Gly-32 are moved out of hydrogen-bonding distance in the triple-stranded antiparallel β-sheet. Analyses of the protein–water hydrogen bonding were also carried out and are compared with results from previous simulations and NMR experiments.

The rational design or modification of biomolecules, including the development of selective inhibitors for enzymes, requires detailed knowledge of the structure, dynamics, and corresponding energetics. Importantly, the continuous improvement of crystallographic techniques has made possible the precise determination of the structures of many proteins, as reflected in the more than 300 entries now deposited in the Brookhaven Protein Data Bank.[1] However, the structures obtained in this fashion represent an ordered crystalline state, while biological processes normally occur in solution. Furthermore, the data obtained from crystal structures are static in nature, although some dynamic information can be obtained from the temperature (*B*) factors.[2] Recent advances in nuclear magnetic resonance spectroscopy (NMR), particularly the use of 2-D nuclear Overhauser effects, have been very valuable, since sets of distance constraints are obtained that can be transformed into three-dimensional structures of proteins in solution.[3] Although this methodology allows the direct observation of proteins in their native solution state, the structures obtained reflect conformational averaging and are not unique solutions for the data sets. Nevertheless, the combination of the experimental structural results with molecular dynamics (MD) calculations is proving to be a powerful approach to the detailed characterization of the structure, dynamics, and energetics of proteins.[4]

Since the pioneering MD simulation of bovine pancreatic trypsin inhibitor (BPTI) in vacuo,[5] there have been numerous molecular dynamics calculations of proteins.[4] However, the aqueous medium has rarely been represented in molecular detail. Some exceptions are for BPTI,[6-8] avian pancreatic polypeptide hormone (APP),[9] and parvalbumin[10] in aqueous solution and BPTI in its full crystalline environment.[11] Other recent calculations have been more focused toward the modeling of enzyme–inhibitor complexes in water, including trypsin–benzamidine[12] and thermolysin–phosphonamidate.[13]

With the exceptions of the parvalbumin[10] and one of the BPTI calculations,[8] all of these simulations were run for very short times. Total times including the equilibrium periods were 15–30 ps for some of the BPTI and the APP simulations and 45 ps for the trypsin–benzamidine complex, as compared to 106 and 210 ps for the parvalbumin and the longest of the BPTI calculations, respectively. It is not clear that the shorter simulation times allow the systems to achieve equilibrium and to remove the biases from the starting conformation, typically obtained from a crystal structure. In this setting, the present study was undertaken to follow a molecular dynamics simulation for a protein in water for a long enough time to assess the convergence issue, to further test the performance of the OPLS force field,[14] and to obtain insights on protein hydration and possible differences between the solution and crystal structures. The protein selected for the study was the third domain of silver pheasant ovomucoid (OMSVP3).[15]

Ovomucoids make up about 10% of the protein in avian egg whites, in which they are the dominant inhibitors of serine proteases. They are members of the Kazal family of pancreatic secretory inhibitors, which are generally important in controlling the premature activation of pancreatic zymogens. The complete ovomucoid consists of three homologous, tandem domains, each

(1) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
(2) Petsko, G. A.; Ringe, D. *Annu. Rev. Biophys. Bioeng.* **1984**, *13*, 331. Fraunfelder, H.; Hartmann, H.; Karplus, M.; Kuntz, I. D., Jr.; Kuriyan, J.; Parak, F.; Petsko, G. A.; Ringe, D.; Tilton, R. A., Jr.; Connolly, M. L.; Max, N. *Biochemistry* **1987**, *26*, 254.
(3) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986.
(4) For recent reviews, see: (a) Karplus, M.; McCammon, J. A. *CRC Crit. Rev. Biochem.* **1986**, *9*, 293. (b) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, England, 1987.
(5) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.
(6) van Gunsteren, W. F.; Berendsen, H. J. C. *J. Mol. Biol.* **1984**, *176*, 559.
(7) Ghosh, I.; McCammon, J. A. *J. Phys. Chem.* **1987**, *91*, 4878.
(8) Levitt, M.; Sharon, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7557.
(9) Krüger, P.; Straßburger, W.; Wollmer, A.; van Gunsteren, W. F. *Eur. Biophys. J.* **1985**, *13*, 77.
(10) (a) Ahlström, P.; Teleman, O.; Jönsson, B.; Forsén, S. *J. Am. Chem. Soc.* **1987**, *109*, 1541. (b) Ahlström, P.; Teleman, O.; Jönsson, B. *J. Am. Chem. Soc.* **1988**, *110*, 4198.
(11) (a) van Gunsteren, W. F.; Karplus, M. *Biochemistry* **1982**, *21*, 2259. (b) Swaminathan, S.; Ichiye, T.; van Gunsteren, W. F.; Karplus, M. *Biochemistry* **1982**, *21*, 5230. (c) van Gunsteren, W. F.; Berendsen, H. J. C.; Hermans, J.; Hol, W. G. J.; Postma, J. P. M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 4315.
(12) Wong, C. F.; McCammon, J. A. *Isr. J. Chem.* **1986**, *27*, 211; *J. Am. Chem. Soc.* **1986**, *108*, 3830.
(13) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* **1987**, *235*, 574.
(14) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(15) A preliminary report on this work was provided in the Proceedings of the 1988 Nobel Symposium: Jorgensen, W. L.; Tirado-Rives, J. *Chem. Scr.* **1989**, *29A*, 191.

0002-7863/90/1512-2773$02.50/0

of which may inhibit a protease. The third domain contains 56 amino acid residues that can be detached by controlled proteolysis. These fragments are typically at least as active as the full ovomucoid.

Sequences have now been determined for ovomucoid third domains from over 100 avian species by Laskowski and coworkers.[16] They have also determined association constants for many of these inhibitors with α-chymotrypsin (AC), elastase (HLE), subtilisin, and *Streptomyces griseus* proteases A and B (SGPA and SGPB).[17,18] Many of the sequences differ by only one or a few residue changes, so an unusually complete structure/activity data base is being constructed.

Besides the association and hydrolytic constants, some structural data are also available. Crystal structures have been determined for complexes of turkey ovomucoid third domain with SGPB,[19] HLE,[20] and AC,[21] for two isolated third domains, silver pheasant[22] and Japanese quail,[23] and for their corresponding hydrolyzed forms.[24] However, structural data in solution are limited to two studies by 2-D NMR, specifically, for turkey ovomucoid third domain in both its native[25] and hydrolyzed forms.[26]

The typical structure of an ovomucoid third domain contains three disulfide bridges, a 10–11 residue long α-helix, and a triple-stranded antiparallel β-sheet. The combination of small size, a large body of experimental data, and interesting structure makes the ovomucoids unusually attractive for a series of molecular dynamics investigations. Such simulations could provide detailed structural and thermodynamic information, which would be of great assistance in the interpretation of the biophysical data and the development of selective inhibitors. The third domain of silver pheasant ovomucoid was chosen for this initial study owing to its greater sequence homology with other ovomucoids and the better resolution of its crystal structure than that of Japanese quail.

Computational Procedure

The entire simulation was conducted on Sun-4 computers using the AMBER 3.0 program,[27] with minor local modifications to improve its use of the UNIX environment. The OPLS nonbonded parameters[14] were used for the protein atoms in conjuction with the TIP3P model for water.[28] As specified in the OPLS model, the dielectric constant was kept fixed at 1.0, and the scaling factors for the 1,4-nonbonded interactions were 8.0 for the Lennard-Jones and 2.0 for the electrostatic interactions.[14] The energetics for angle bending and torsional motion were described with the AMBER united-atom force field.[27] During the simulation, all bond lengths and the H–H distances in water were kept constant by using the SHAKE algorithm[29] with a tolerance of 0.0004 Å, which allowed the use of a time step of 2 fs. The temperature and pressure were kept constant at 298 K and 1 bar (0.987 atm). A nonbonded pair list was used to accelerate the calculations and was updated every 10 steps. This list was generated by using a residue-based cutoff (9 Å) to avoid splitting dipoles. All the calculations utilized periodic boundary conditions to avoid edge effects.

The initial coordinates for the third domain of silver pheasant ovomucoid were obtained from the crystal structure[22] deposited in the Brookhaven Protein Data Bank.[1,30] The protein molecule, without any of the crystallographically located water molecules, was centered in a rectangular box of water obtained by periodic translations in the x, y, and z directions of a cube of water previously equilibrated via Monte Carlo calculations. Any water molecule closer than 1.5 Å to any protein atom or farther away than 6 Å from the closest protein atom in any one Cartesian direction was then deleted to give an initial system containing the solute plus 1721 water molecules (5676 total atoms) in a rectangular box of dimensions 43.8 × 42.0 × 34.3 Å.

The initial preparation of the system consisted of 100 steps of steepest descent energy minimization, followed by a short (1 ps) constant volume molecular dynamics run. The resulting structure was then used as the starting point for the MD simulation at constant temperature and pressure. Initial atomic velocities were assigned from a Maxwellian distribution corresponding to a temperature of 298 K. The values of the potential energy, RMS deviation from the crystal structure, and volume were monitored continuously in order to follow the equilibration of the system. A total time of 100 ps was covered in the simulation, during which the coordinates, velocities, and energies were saved every 50 time steps (0.1 ps) for further analysis.

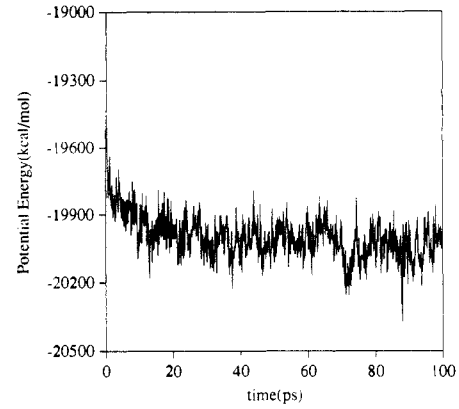

Figure 1. Potential energy variation during the course of the MD simulation.

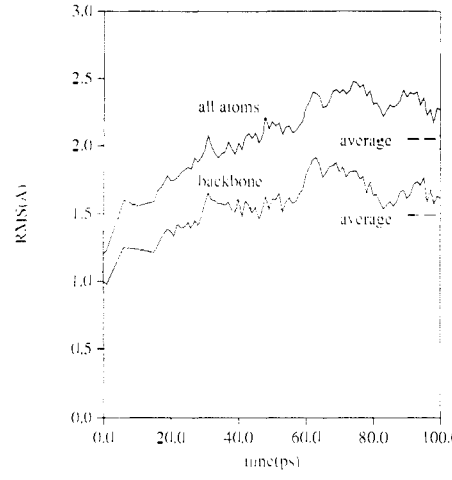

Figure 2. RMS deviations between the instantaneous computed structure and the crystal structure for all residues as a function of simulation time.

Results

Convergence Behavior. The potential energy and RMS deviations of the main-chain atoms (N, C$^\alpha$, C, and O) of residues 8–56 for each instantaneous structure are plotted in Figures 1 and 2, respectively, as a function of simulation time. The first seven

(16) Laskowski, M., Jr.; Kato, I.; Ardelt, W.; Cook, J.; Denton, A.; Empie, M. W.; Kohr, W. J.; Park, S. J.; Parks, K.; Schatzley, B. L.; Schoenberger, O. L.; Tashiro, M.; Vichot, G.; Whatley, H. E.; Wieczorek, A.; Wieczorek, M. *Biochemistry* **1987**, *26*, 202. Kato, I.; Kohr, W. J.; Laskowski, M., Jr. *Biochemistry* **1987**, *26*, 193.

(17) Empie, M. W.; Laskowski, M., Jr. *Biochemistry* **1987**, *21*, 2274.

(18) Laskowski, M., Jr.; Tashiro, M.; Empie, M. W.; Park, S. J.; Kato, I.; Ardelt, W.; Wieczorek, M. In *Proteinase Inhibitors: Medical and Biological Aspects*; Katunuma, N., Umezawa, H., Hozer, H., Eds.; Japan Scientific Societies Press: Tokyo, 1983; p 55.

(19) Fujinaga, M.; Read, R.; Sielecki, A. R.; Ardelt, W.; Laskowski, M., Jr.; James, M. N. G. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4868. Read, R.; Fujinaga, M.; Sielecki, A. R.; James, M. N. G. *Biochemistry* **1983**, *22*, 4420.

(20) Bode, W.; Wei, A.-Z.; Huber, R.; Meyer, E.; Travis, J.; Neumann, S. *EMBO J.* **1986**, *10*, 2453.

(21) Read, R.; Fujinaga, M.; Sielecki, A. R.; Ardelt, W.; Laskowski, M., Jr. *Acta Crystallogr., Sect. A, Suppl.* **1984**, *40*, C-50.

(22) Bode, W.; Epp, O.; Huber, O.; Laskowski, M., Jr.; Ardelt, W. *Eur. J. Biochem.* **1985**, *147*, 387.

(23) Weber, E.; Papamokos, E.; Bode, W.; Huber, R.; Kato, I.; Laskowski, M., Jr. *J. Mol. Biol.* **1981**, *149*, 109. Papamokos, E.; Weber, E.; Bode, W.; Huber, R.; Empie, M. W.; Kato, I.; Laskowski, M., Jr. *J. Mol. Biol.* **1982**, *158*, 515.

(24) Musil, D.; Bode, W.; Mayr, I.; Huber, R.; Laskowski, M., Jr.; Lin, T.-Y.; Ardelt, W. Unpublished results.

(25) Robertson, A.; Westler, W. M.; Markley, J. L. *Biochemistry* **1988**, *27*, 2519.

(26) Rhyu, G. I.; Markley, J. L. *Biochemistry* **1988**, *27*, 2529.

(27) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. J. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(28) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *70*, 926.

(29) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.

(30) Entry 2OVO, version of Nov 8, 1985.

residues were excluded from the RMS deviations, since they form a pendant tail that is less well resolved in the crystal structure.[22] The RMS deviations are typically 0.2 Å higher when these residues are included. The system appears to have reached an equilibrium state after 30–40 ps, though the ultimate longevity of this state could only be determined in a much longer run. After this initial period, the simulation was continued to 100 ps in order to acquire enough data for averaging and analysis. In the 210-ps simulation for the other protease inhibitor, BPTI, in water, an equilibrium state was achieved after ca. 50 ps.[8]

Protein Structure and Dynamics. (A) Comparison with the Crystal Structure. The calculated mean structure of the protein in solution, obtained by directly averaging the Cartesian coordinates of all the saved structures from 30 to 100 ps, shows an RMS deviation from the crystal structure of 1.28 Å for the backbone atoms of residues 8–56 and 1.49 Å for all residues. Results from previous simulations in water include 1.05 Å for the C^{α} of APP, averaged from 5 to 15 ps,[9] and 0.77 Å for all the backbone atoms of residues 1–56 of BPTI, averaged from 105 to 210 ps.[8] Deviations from simulations in vacuo are typically larger by at least a factor of 2.[8,11]

The value of the RMS deviation of the instantaneous structure at 100 ps from the crystal structure is 1.43 Å for the backbone atoms of residues 8–56 and 1.61 Å for all residues. These values are in the same range as those obtained in previous, shorter simulations in water, e.g., 1.5 Å for only the C^{α} atoms of BPTI after 20 ps[6] and 1.72 Å for the C^{α} atoms of the trypsin–benzamidine complex at 45 ps.[12] Clearly, the AMBER/OPLS force field is providing a comparatively reasonable representation of the protein. An approximate base line for the RMS deviations can be deduced from the comparison of coordinates from proteins whose crystal structures have been solved in different crystalline forms or that have different molecules in the asymmetric unit. For five such cases, Chothia and Lesk obtained RMS deviations ranging from 0.25 to 0.40 Å, with a mean of 0.33 Å.[31]

A comparative plot of the backbone atoms of the crystal structure with the instantaneous structure at the end of the MD run is given in Figure 3. It can be seen that the tertiary structure is well preserved in the simulation. However, several striking differences with the crystal structure are revealed by more detailed graphical analyses:

(1) In the crystal structure, the amino acid side chains are mostly folded onto the surface of the protein, while in the solution simulation they extend more into the solvent. This effect likely reflects the lower water content in the crystal and the removal of the interprotein interactions. It also allows polar side chains to be better stabilized by hydration.

(2) The last residue of the α-helix in the crystal, Ser-44, does not form the required hydrogen bond with Ala-40 to be part of the helix in solution. This terminal hydrogen bond is also absent in the crystal structures of isolated Japanese quail[23] and hydrolyzed silver pheasant[24] ovomucoids and in the complexes of turkey ovomucoid with α-chymotrypsin[21] and *S. griseus* protease B.[19]

(3) As discussed in the next section, some of the residues of the outermost strand of the β-sheet (Lys-29, Thr-30, Tyr-31, and Gly-32) are tilted out of the plane of the other two strands in the simulation. This displacement inhibits the formation of the interstrand hydrogen bonds and yields greater hydrogen bonding between these residues and water molecules.

(4) The hydrogen bond in the crystal structure between the side chains of Glu-19 and Thr-17, which surround the scissile peptide bond between Met-18 and Glu-19, is not found in the simulation results. Instead, it is replaced by a hydrogen bond between the guanidinium fragment of Arg-21 and the carboxylate in Glu-19, while the hydroxyl group of Thr-17 is hydrogen-bonded to solvent molecules. This effect appears to be related to the loss of interprotein interactions. In the crystal lattice, the charged side charged side chain of Arg-21 is close to the carbonyl oxygens of Glu-10 and Pro-12 from a protein molecule in a neighboring unit cell. Interestingly, the same interprotein contacts were found in the crystal structure of Japanese quail ovomucoid, which has a very different crystalline form.[23,32] In addition, the Glu-19 carboxylate group in the crystal is hydrogen-bonded to the side-chain ammonium group of Lys-13 in the same neighboring cell. Figure 4 compares the different interactions in this region for the crystal structure and the instantaneous structure at 80 ps.

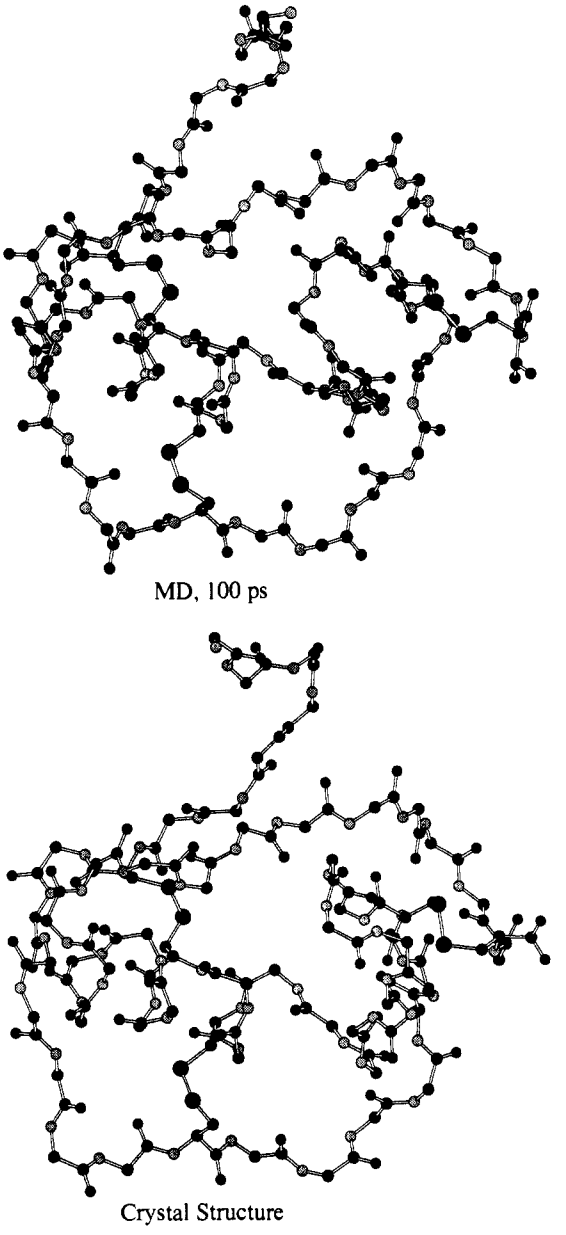

Figure 3. Comparison of backbone atoms in the instantaneous structure at the end of the simulation (t = 100 ps) and in the crystal structure.

The overall conformation of a protein can be expressed in terms of the backbone torsional angles ϕ_i (C_{i-1}–N_i–C_i^{α}–C_i) and ψ_i (N_i–C_i^{α}–C_i–N_{i+1}). The average values for these angles during the 30–100-ps period are compared with the corresponding values for the crystal structure in Figure 5. Consistent with the statements above, the biggest differences are found in the region around the C-terminus of the α-helix (Glu-43, Ser-44, and Asn-45), where the middle residue is twisted away from the helix, and in the region near the turn connecting the central and outer strands of the β-sheet (Ser-26, Asp-27, and Asn-28). Some of the other differences are of little consequence to the overall conformation, since compensation occurs when there are differences of opposing signs in an angle ψ_i and the subsequent ϕ_{i+1}. This is the case for Met-18 and Glu-19 and for the Thr-47, Leu-48, and Thr-49 region. The

(31) Chothia, C.; Lesk, A. M. *EMBO J.* **1986**, *5*, 823.

(32) Silver pheasant ovomucoid crystallizes in the *C*2 space group, while for Japanese quail, the crystals belong in the tetragonal $P4_12$ space group.

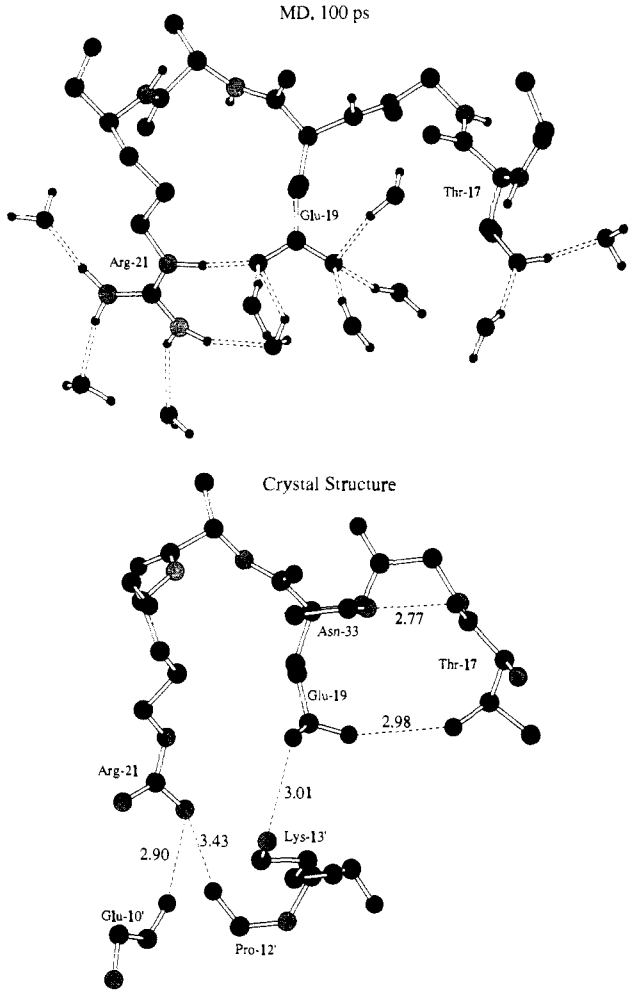

Figure 4. Hydrogen bonds near the scissile bond in the instantaneous structure at the end of the simulation (t = 100 ps) and in the crystal structure.

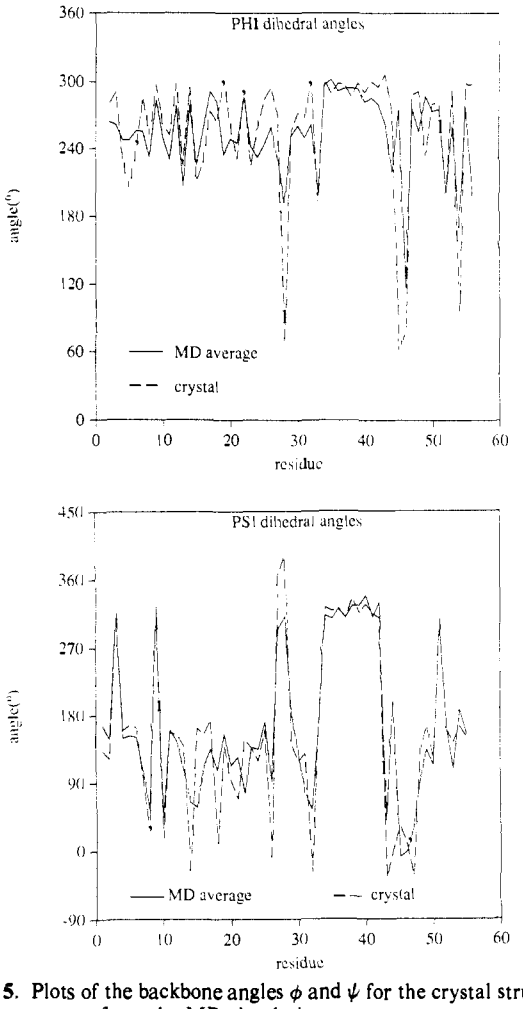

Figure 5. Plots of the backbone angles ϕ and ψ for the crystal structure and the averages from the MD simulation.

RMS differences from the crystal structure computed for each backbone dihedral angle over the entire protein are 39°, 48°, and 11° for ϕ, ψ, and ω, respectively. These differences are in the same range as the results from previous simulations, e.g., 31°, 37°, and 8° for ϕ, ψ, and ω in trypsin[12] and 26° and 33° for ϕ and ψ for BPTI in "van der Waals water".[11a]

A more global impression of the overall conformation of the protein can be obtained from the plots of ϕ vs ψ (Ramachandran maps) in Figure 6. The general trend observed in the maps is that residues in the α-helix stay close to the crystallographically observed angles, while those outside the helix are shifted toward values more typical of the C_5 and C_7^{eq} conformations. This displacement is consistent with the results of our previous energy-minimization studies on the conformations of *N*-acetylglycine-*N*-methylamide and *N*-acetylalanine-*N*-methylamide, which showed that these conformations are lower in energy than the corresponding α, α_R, or α_L alternatives.[14]

(B) Comparison with NMR Data. The NMR data in solution on native and hydrolyzed turkey ovomucoid should be relevant to the present case, since turkey and silver pheasant ovomucoids differ by only one residue (18: Met/Leu). In general, the main effort in the NMR investigations was devoted to the complete assignment of resonances, and only qualitative information regarding secondary structure was obtained.[25,26]

Although the principal conclusion of these studies was that the solution structure had to be very similar to that in the crystal, some details were explored further. In particular, the NOESY connectivities that establish the antiparallel triple-stranded β-sheet were given. Figure 7 provides a comparison of the segments comprising the β-sheet in the crystal and in the mean calculated solution structure, obtained by direct averaging of the Cartesian coordinates from 30 to 100 ps in the MD simulation. The interproton distances given with the crystal fragments are measured in the crystal[33] but correspond to observed NOE pairs in the solution NMR. Only the distances that show significant deviations from the MD averages are given. As alluded to above, the largest differences are found in the separation between the central and the outermost strands of the β-sheet. The two distances of 6–7 Å calculated in the simulation are beyond the normally accepted range for NOE detection. A problem with the force field or the preparation of the system could be indicated. However, it is also possible that the protein is exploring the expanded phase of a low-frequency "breathing mode", and the separations may decrease at later times in the simulation.[34]

(C) Fluctuations. The overall mobility of the different atoms during an MD simulation can be expressed as their RMS fluctuations, $\langle \Delta r^2 \rangle^{1/2} = \langle (r_i - \langle r_i \rangle)^2 \rangle^{1/2}$, computed over the averaging period. An examination of the accumulated average atomic fluctuations as the time span of their evaluation increases shows, as expected, a monotonic increase for about 40 ps that levels toward a plateau after ca. 50 ps. These fluctuations can be compared to the average movement observed in the crystallographic determination as expressed in the B thermal factors. Crystallographic RMS fluctuations can then be derived via the relation $\langle \Delta r^2 \rangle^{1/2} = (3B/8\pi^2)^{1/2}$ from the B-values.[35] Figure 8 compares the RMS fluctuations calculated during the last 25 ps of the simulation with the fluctuations derived from the thermal factors.[22]

(33) Any missing hydrogen atoms in the X-ray and the MD structures were added by using the SYBYL program from TRIPOS Associates.
(34) Suezaki, Y.; Go, N. *Int. J. Pept. Protein Res.* **1975**, *7*, 333.
(35) Willis, B. T. M.; Pryor, A. W. *Thermal Vibrations in Crystallography*; Cambridge University Press: Cambridge, England, 1975.

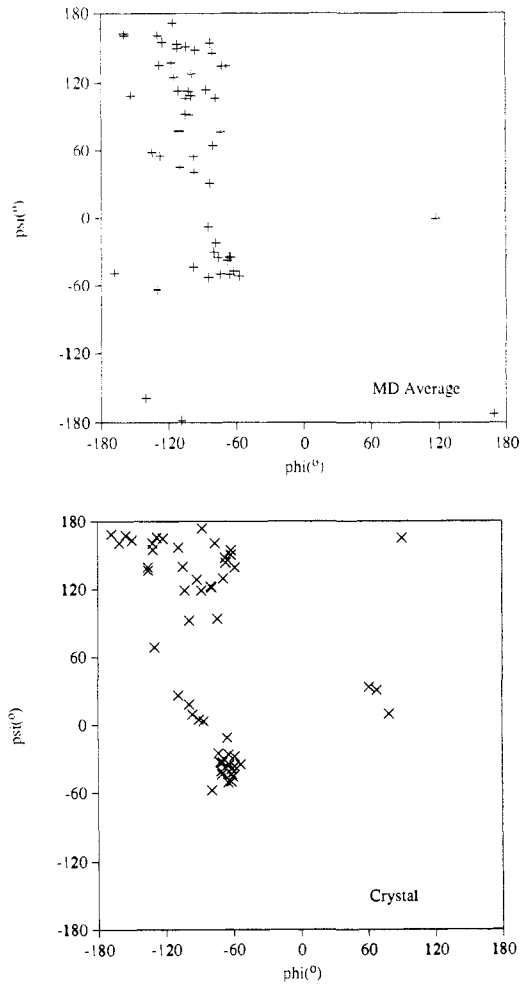

Figure 6. Ramachandran maps for the MD results and the crystal structure.

Table I. Average RMS Fluctuations (Å) for Different Atom Types

	averaging time, ps		
atoms	30–50	30–70	30–100
C^{α}	0.73	1.16	1.32
C^{β}	0.82	1.24	1.40
C^{γ}	0.90	1.30	1.50
C^{δ}	0.92	1.30	1.45
all protein	0.86	1.27	1.44
O^{water}	2.76	3.96	5.17

It can be seen from the plot that qualitative agreement in the width and location of the largest peaks is obtained. The most noticeable difference is the comparatively high fluctuations for the N-terminus in the simulation, which decrease until Cys-8 is reached. The first seven residues form a pendant chain whose increased mobility in solution can be attributed to the formation of a four-stranded β-channel with the N-terminal segments of neighboring molecules in the crystal.[22] The remaining qualitative differences reflect greater motion in the simulation for the side chains of Glu-10, Met-18, Asp-27, Lys-29, Glu-43, and Asn-45, which are more exposed to the solvent in the simulation than in the crystal structure.

The RMS fluctuations for the different atoms are correlated to their distances from the backbone of the protein. Table I lists the average RMS fluctuations for different carbon atoms for three averaging periods. The values obtained for the C^{δ} atoms have larger statistical uncertainties, since there are only 30 of these atoms in the protein. The fluctuations calculated in the initial 20 ps of the averaging period are comparable to those obtained in the simulations of aqueous trypsin–benzamidine during 19 ps (C^{α} = 0.52, C^{β} = 0.59, C^{γ} = 0.69, C^{δ} = 0.70, all = 0.68),[12] BPTI in "van der Waals water" for 25 ps (C^{α} = 0.54, C^{β} = 0.69, C^{γ} = 0.89, C^{δ} = 1.15, all = 0.78),[11a] and aqueous APP for 10 ps (C^{α} = 0.53, C^{β} = 0.62, C^{γ} = 0.73, C^{δ} = 0.79, all = 0.75).[9] However, the fluctuations calculated over the full averaging period of 70 ps are considerably larger than those reported by Levitt and Sharon for the final 105 ps of residues 2–56 of BPTI (C^{α} = 0.42, C^{β} = 0.50, C^{γ} = 0.54, C^{δ} = 0.67).[8] Greater experience is needed to ascertain if these differences are associated more with the proteins, the force fields, or the details of the simulation procedures, though further comment on this issue is made below.

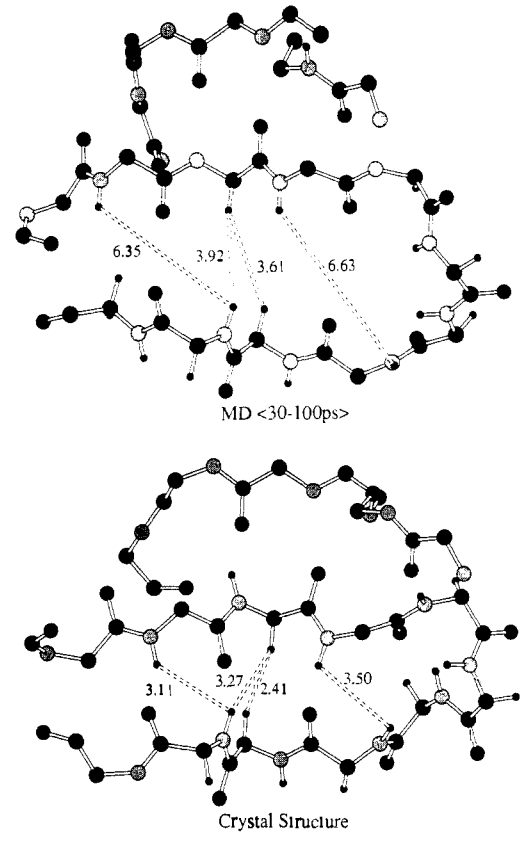

Figure 7. Backbone atoms of the β-sheet segments for the average MD structure and in the crsytal structure. Some interproton distances are shown.

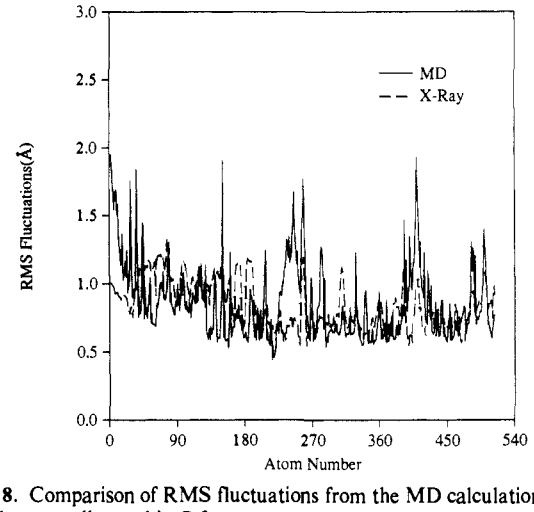

Figure 8. Comparison of RMS fluctuations from the MD calculation and from the crsytallographic *B*-factors.

(D) Hydrogen Bonding. Analyses of the intraprotein hydrogen bonding were carried out on the 700 coordinate sets saved during the last 70 ps of the MD simulation. The criteria used to define a hydrogen bond were purely geometric. A list of all potential donors and acceptors (hydrogens attached to heteroatoms and the heteroatoms themselves) was generated at the beginning of the analysis. For each coordinate set, every potential donor–acceptor

Table II. Protein–Protein Hydrogen Bonds and Their Frequencies

donor		acceptor		crystal	MD
atom	residue	atom	residue	distance, Å	frequency, %
		Backbone to Backbone			
N	Tyr-11	O	Cys-8	3.4	22.6
N	Arg-21	O	Gly-32	2.7	99.6
N	Leu-23	O	Tyr-31	2.9	
N	Cys-24	O	His-52	2.8	99.4
N	Gly-25	O	Lys-29	2.8	
N	Ser-26	O	Thr-49	2.8	96.6
N	Asn-28	O	Gly-25	2.9	
N	Tyr-31	O	Leu-23	2.8	
N	Phe-37	O	Asn-33	3.0	94.1
N	Cys-38	O	Lys-34	2.8	100.0
N	Asn-39	O	Cys-35	3.0	98.9
N	Ala-40	O	Asn-36	3.1	52.3
N	Val-41	O	Phe-37	2.9	70.6
N	Val-42	O	Cys-38	3.1	98.6
N	Glu-43	O	Asn-39	2.9	85.3
N	Ser-44	O	Ala-40	2.9	0.3
N	Asn-45	O	Val-42	2.9	
N	Gly-46	O	Val-41	2.8	8.0
N	Thr-47	O	Ser-44	3.2	
N	Ser-51	O	Cys-24	2.9	55.3
N	His-52	O	Cys-24	3.2	77.7
N	Cly-54	O	Pro-22	3.1	53.4
		Backbone to Side Chain			
N	Lys-13	OD1	Asn-39	3.1	99.0
N	Lys-29	OD2	Asp-27	3.1	44.1
N	Asn-36	OD1	Asn-33	3.0	99.9
N	Leu-48	OG	Ser-44	3.0	
N	Thr-49	OG	Ser-26	3.0	
N	Cys-56	OG1	Thr-30	3.0	
		Side Chain to Backbone			
OG	Ser-26	O	Thr-49	3.2	1.4
NZ	Lys-34	O	Asp-7	3.1	
ND2	Asn-33	O	Thr-17	2.8	0.1
ND2	Asn-33	O	Glu-19	3.0	
ND2	Asn-39	O	Lys-13	3.1	82.9
OG	Ser-44	O	Ala-40	3.1	64.7
OG	Ser-44	O	Val-41	2.9	13.7
OG1	Thr-47	O	Ser-44	2.9	
ND1	His-52	O	Phe-53	2.6	23.7
		Side Chain to Side Chain			
OG	Ser-9	OD1	Asp-7	2.5	79.0[a]
OG1	Thr-17	OE2	Glu-19	3.0	
OG	Ser-26	OG1	Thr-49	3.2	
OEH	Tyr-31	OD1	Asp-27	2.7	80.1[b]
NZ	Lys-55	OEH	Tyr-20	3.2	

[a] The reported percentage corresponds to the population of hydrogen bonds to either of the two oxygens in the carboxylate acceptor. The hydrogen bonds to each oxygen have occurrences of 50.4 and 32.9%. [b] The reported percentage corresponds to the population of hydrogen bonds to either of the two oxygens in the carboxylate acceptor. The hydrogen bonds to each oxygen have occurrences of 52.7 and 32.0%.

pair was tested and was considered to form a hydrogen bond if the hydrogen to acceptor distance was within 2.5 Å and the donor–hydrogen–acceptor angle was between 120° and 180°.[36]

During the averaging period, 54 protein–protein hydrogen bonds were detected with occurrences of at least 10%. Table II contains a list of all the intraprotein hydrogen bonds detected in the crystal structure and the frequency of their occurrence in the MD simulation. Out of the 42 hydrogen bonds in the crystal, 23 were detected at least 10% of the time, and 19 of these had frequencies of more than 50%. The hydrogen bonds absent in the simulation are substituted by hydrogen bonds either to different solute atoms or to solvent molecules.

The backbone to backbone hydrogen bonds are the most important for the secondary structure and are for the most part

(36) Repetition of the present hydrogen-bonding analysis using the criteria proposed by Levitt and Sharon[8] (acceptor···H < 2.6 Å and acceptor···donor–H < 35°) produced results within 3% of those reported here.

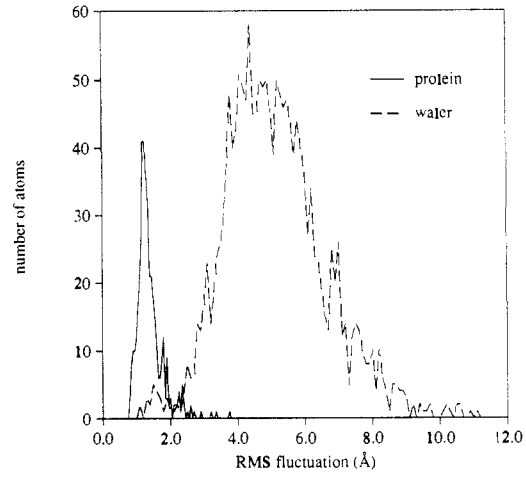

Figure 9. Distribution of RMS fluctuations for heavy atoms of the protein and water oxygens from the MD simulation.

conserved in the MD simulation. This is also reasonable, since the backbone atoms show smaller RMS deviations from the crystal structure than do the side chains. A total of 27 backbone hydrogen bonds were observed during the period from 30 to 100 ps for more than 10% of the structures, with a mean occurrence of 64%. Among these, 14 of the 22 hydrogen bonds detected in the crystal structure are conserved at the 10% or greater level. Some of the previous MD simulations have reported conservation of 13 out of 16 crystallographic backbone hydrogen bonds in the simulation of BPTI in a truncated octahedron of water,[6] 16 out of 20 in the APP simulation,[9] and 20 out of 21 in Levitt and Sharon's simulation of BPTI in water with similar hydrogen-bond criteria.[8,36] The latter authors also reported a mean occurrence of 68% for the backbone hydrogen bonds during the simulation. Since BPTI and OMSVP3 have similar size and secondary structural elements, the smaller fluctuations and greater hydrogen-bond conservation in the study of Levitt and Sharon suggest that the Lifson force field is stiffer than the OPLS/AMBER model and keeps the protein closer to the starting crystal structure. Another indication of the relative rigidity of their force field may be found in the fluctuations calculated during their MD simulation in vacuo, which, although larger than their corresponding values in solution, are still smaller than the fluctuations observed in the crystal. This is not surprising, since the Lifson force field was specifically designed to reproduce crystal structures.[37]

The backbone hydrogen bonds that are absent from the present solution calculation fall in two major groups: those for residues 44–40, 45–42, and 47–44, formed in the crystal by the last segment of the α-helix and the turn needed to connect back to the β-sheet, and the hydrogen bonds formed in the crystal by the third strand of the β-sheet, namely, 23–31, 25–29, 28–25, and 31–23. These structural features, as mentioned before, are not present in the structure calculated in solution.

Solvent Structure and Dynamics. In the course of the molecular dynamics simulation, the solvent exhibits, as expected, greater mobility than the protein. The average RMS fluctuation of the protein atoms from 30 to 100 ps is 1.44 Å (σ = 0.42), while the corresponding quantity for water molecules is 5.17 Å (σ = 1.48). Distributions of the RMS fluctuations for all the heavy atoms can be seen in Figure 9. It is notable that there is some overlap in the range of fluctuations for protein and water atoms. That is, some of the water molecules exhibit behavior more characteristic of the protein than of the solvent, as discussed further below.

(A) Self-Diffusion. Diffusion coefficients were calculated from the squared displacements by using the Einstein relation

$$\lim_{\Delta t \to \infty} \langle [R(t + \Delta t) - R(t)]^2 \rangle = 6D\Delta t$$

as the slope of the mean square displacements vs Δt for periods

(37) Lifson, S.; Hagler, A. T.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5111. Hagler, A. T.; Dauber, P. *Acc. Chem. Res.* **1980**, *13*, 105.

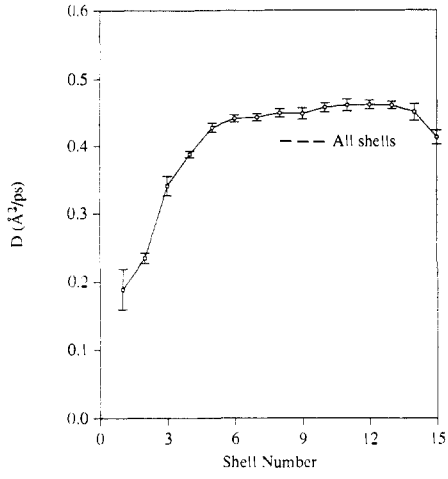

Figure 10. Calculated diffusion coefficients for water molecules as a function of their starting distance to the closest protein atom.

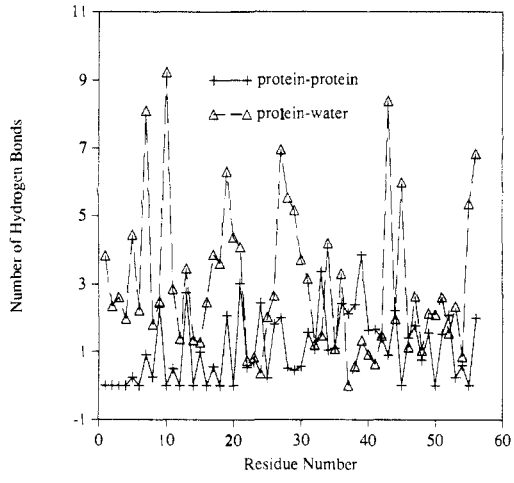

Figure 11. Average number of computed protein–protein and protein–water hydrogen bonds for each residue.

of up to 20 ps and averaged over all choices of initial time t. In order to obtain a more complete idea of the dynamic behavior of the solvent, the squared displacements were calculated for each water molecule and then averaged separately as a function of the distance from the closest protein atom at the beginning of each time span, Δt, for 1-Å "shells". There is some crossing between different shells at the longer times, making the separation less than perfect. Nevertheless, the results in Figure 10 are reasonable. The diffusion coefficient is low at small distances from the protein surface, 0.19 $Å^2 \cdot ps^{-1}$ for the first shell (water molecules less than 2 Å from the protein), and increases rapidly to a plateau of approximately 0.45 $Å^2 \cdot ps^{-1}$ that spans from 5 to 14 Å. The value obtained for the last shell (0.41 $Å^2 \cdot ps^{-1}$) has a larger statistical uncertainty due to the reduced number of solvent molecules located in this region. The average diffusion coefficient for all the water molecules is 0.41 $Å^2 \cdot ps^{-1}$, which is higher than the experimentally determined value for pure water, 0.23 $Å^2 \cdot ps^{-1}$ at 25 °C.[38]

Previous simulations reported varied behavior for the diffusion coefficient as a function of distance to the protein. In the simulation of the trypsin–benzamidine complex in SPC water, Wong and McCammon found an almost linear increase of the diffusion coefficient from 0.15 $Å^2 \cdot ps^{-1}$ close to the protein to a value slightly higher than 0.5 for water molecules 15 Å away.[12] The diffusion coefficient calculated for bulk SPC water at 25 °C is 0.36 $Å^2 \cdot ps^{-1}$.[39] (It may be noted that the TIP3P and the SPC models are nearly identical.[28]) For the parvalbumin simulation with a modified SPC potential, Ahlström et al. obtained a high diffusion coefficient of approximately 1.0 $Å^2 \cdot ps^{-1}$ close to the protein surface, followed by an almost linear decrease after 5–7 Å until the bulk value of their water model, 0.61 $Å^2 \cdot ps^{-1}$, was reached at 15 Å.[10,40] Levitt

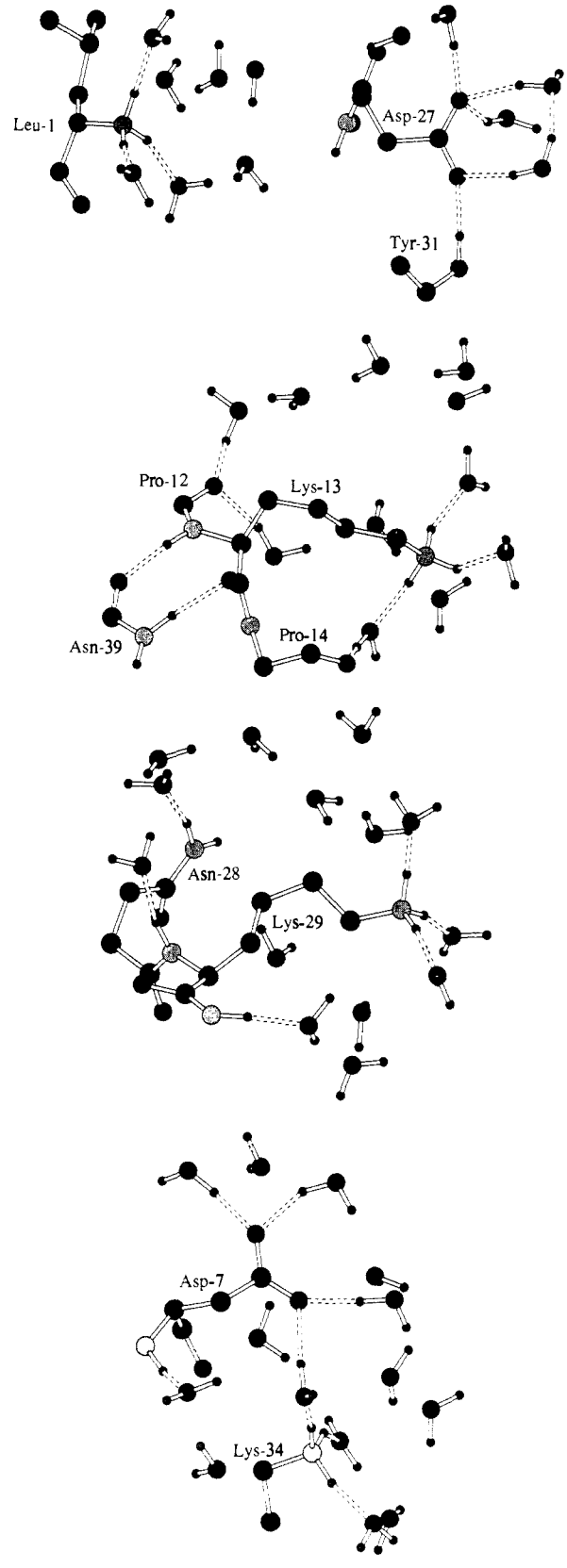

Figure 12. Hydration shells around some carboxylate and ammonium groups for the instantaneous structure at 80 ps.

(38) Mills, R. *J. Phys. Chem.* **1973**, *77*, 685.
(39) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; p 331.
(40) Teleman, O.; Ahlström, P. *J. Am. Chem. Soc.* **1986**, *108*, 4333.

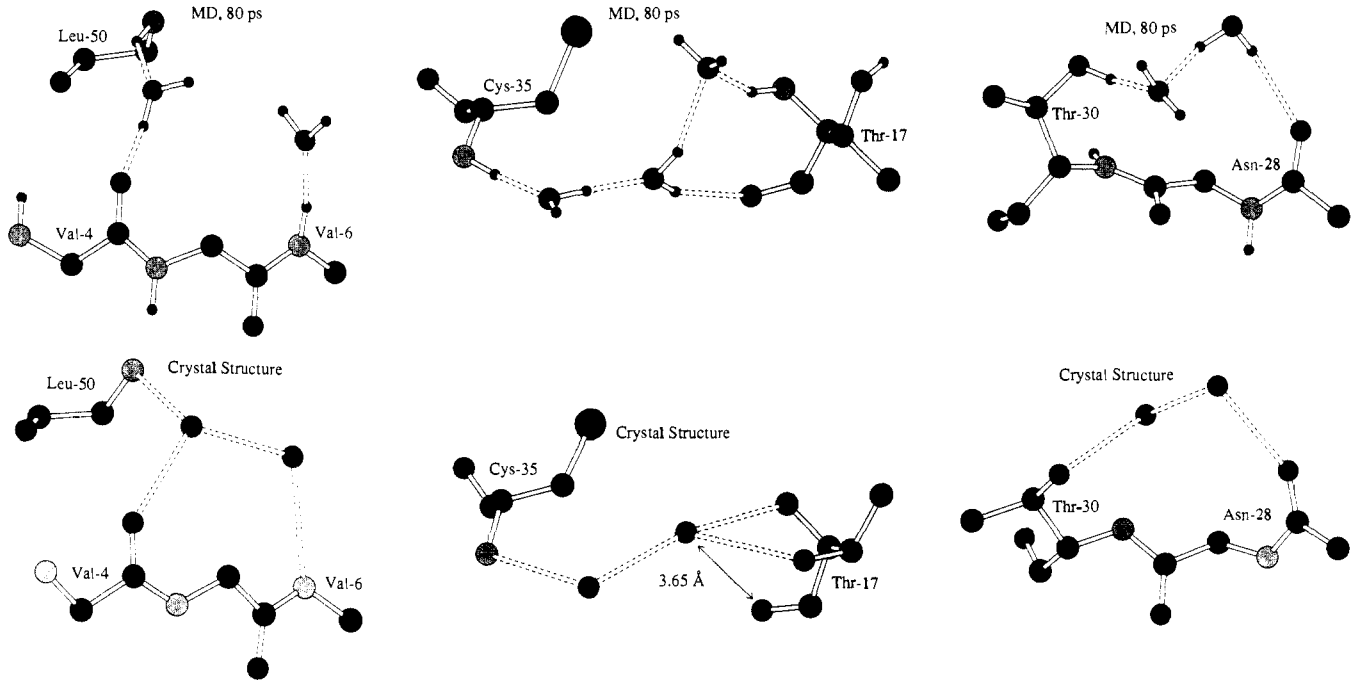

Figure 13. Some low-mobility water molecules from the MD simulation and the corresponding crystallographically determined water positions. The coordinates from the MD simulation represent the instantaneous structure at 80 ps.

and Sharon followed a different analysis scheme for the simulation of BPTI in their F3P model of water.[8] Namely, they divided the water molecules into four broad classes according to their distance from polar or nonpolar protein atoms. In the two classes closest to the protein surface (<3.2 Å from polar atoms and <4.5 Å from nonpolar atoms, respectively), they found reduced values for the diffusion coefficient of 0.10 $Å^2 \cdot ps^{-1}$ and 0.15 $Å^2 \cdot ps^{-1}$, while the water molecules farthest from the protein had a diffusion coefficient of 0.24 $Å^2 \cdot ps^{-1}$. Thus, the results from the trypsin, BPTI, and present studies are qualitatively consistent and indicate significantly reduced water mobility near the protein surface. This seems reasonable, owing to the water structuring by charged and hydrophobic side chains and the loss of motion toward the protein.

(B) Hydrogen Bonding to the Protein. Some of the protein–protein hydrogen bonds that are present in the crystal but absent in the MD simulation have been replaced by protein–water hydrogen bonds. Throughout the simulation, the sum of the number of protein–protein and protein–water hydrogen bonds remains almost constant. The averages for the number of hydrogen bonds for each residue type are given in Table III. It should be added that the total number shows little variability for each particular residue type; e.g., the average total numbers of hydrogen bonds for Ala-2, Ala-3, Ala-15, and Ala-40 are 2.35, 2.62, 2.27, and 2.54, respectively. The average number of hydrogen bonds for each residue is also illustrated in Figure 11. It may be noted that for most of the protein the average number of hydrogen bonds to water is higher than the number of intraprotein hydrogen bonds. This trend is reversed in the region from Phe-37 to Val-42, where a larger number of intraprotein hydrogen bonds are needed to form the α-helix. It is also evident that the highest numbers of hydrogen bonds to water are observed for the residues with charged side chains, Asp-7, Glu-10, Glu-19, Asp-27, Glu-43, and the C-terminal Cys-56.

The numbers of protein–water hydrogen bonds for the residues containing carboxylate and ammonium side chains are consistent with results of previous simulations. In the present MD simulation, lysine residues have an average of 4.55 hydrogen bonds to solvent, of which 1.33 are to the backbone atoms, leaving 3.22 for the side chain. Similarly, Jorgensen and Gao found via Monte Carlo calculations that 3.9 water molecules hydrogen-bond to the methylammonium ion.[41] Aspartate and glutamate residues, after subtraction of the backbone atom–solvent hydrogen bonds, have 5.24 and 6.38 hydrogen bonds to water, in reasonable agreement with the 6.4 water molecules for acetate ion reported by Jorgensen and Gao.[41] Some samples of the hydration shells for these groups are given in Figure 12.

Table III. Average Number of Hydrogen Bonds per Residue Type

residue	abundance	solute–solute	solute–solvent	total	Kuntz hydration no.[a]
Ala	4	0.67	1.79	2.45	1.4 ± 0.5
Arg	1	3.03	4.09	7.12	3.1 ± 1.0
Asn	5	2.03	3.53	5.56	2.0 ± 0.5
Asp	2	1.46	7.54	9.00	8.1 ± 1.0
Cys	5	1.23	1.26	2.49	
Glu	3	0.99	7.98	8.97	8.3 ± 1.0
Gly	4	0.82	1.30	2.12	0.9 ± 0.5
His	1	2.08	1.54	3.62	
Leu	3	0.48	1.33	1.81	
Lys	4	1.06	4.55	5.61	5.0 ± 1.0
Met	1	0.00	3.62	3.62	
Phe	2	1.18	1.17	2.35	
Pro	3	0.18	1.16	1.33	3.1 ± 1.0
Ser	5	1.62	2.84	4.46	
Thr	4	1.11	3.10	4.20	
Tyr	3	0.69	3.47	4.15	5.5 ± 1.0
Val	4	0.77	1.58	2.34	0.9 ± 0.5

[a] See ref 42.

The average number of protein–water hydrogen bonds for each residue type can also be compared to the experimental hydration numbers determined by Kuntz for frozen polypeptide solutions by NMR.[42] The hydration numbers from these experiments are given in the last column of Table III. The accord with the computed solute–solvent numbers is excellent in most cases. The MD results indicate somewhat lessened hydration for proline and tyrosine in the ovomucoid than in the frozen polypeptides, although the hydrogen-bonding analysis performed here would not detect waters associated with the nonpolar regions of these residues. The remaining discrepancy is for asparagine, which shows a higher number of hydrogen bonds (3.5) than expected from the hydration number reported by Kuntz (2.0 ± 0.5).[42] Once the number of protein–water hydrogen bonds to backbone atoms has been subtracted, the side chains form an average of 2.5 hydrogen bonds

(41) Jorgensen, W. L.; Gao, J. *J. Phys. Chem.* **1986**, *90*, 2174.

(42) Kuntz, I. D., Jr. *J. Am. Chem. Soc.* **1971**, *93*, 514.

Table IV. Comparison of Protein–Water Hydrogen Bonds

crystal					MD simulation		
donor		acceptor		distance,	water	fluctuation,	frequency,
atom	residue	atom	residue	Å	no.	Å	%
O	Wat-101	O	Thr-30	3.0	562	2.58	63.4
N	Ala-2	O	Wat-102	2.8	1145	2.76	49.4
N	Ser-5	O	Wat-104	3.3	1014	2.71	86.1
OG1	Thr-49	O	Wat-105	2.7	264	1.68	90.7
O	Wat-108	OE2	Glu-43	3.0	922	3.41	86.9
O	Wat-114	O	Val-4	2.8	849	1.53	85.1
N	Leu-50	O	Wat-114	3.0	849	1.53	94.1
O	Wat-115	O	Glu-43	2.9	1061	3.42	41.7
N	Val-6	O	Wat-119	3.1	880	1.94	92.7
O	Wat-120	O	Thr-47	2.8	1151	3.83	53.4
N	Asp-7	O	Wat-121	3.1	782	3.24	79.7
O	Wat-122	O	Leu-48	2.5	848	1.69	87.1
N	Cys-35	O	Wat-123	3.2	841	1.15	99.4
N	Thr-17	O	Wat-124	2.9	912	2.62	49.6
N	Thr-30	O	Wat-125	2.8	954	2.68	48.3
O	Wat-127	OD1	Asp-27	3.2	1587	2.97	87.4
O	Wat-128	O	Leu-50	2.8	1491	1.71	92.9
O	Wat-129	COO	Cys-56	2.5	659	1.53	92.6
O	Wat-130	O	Gly-25	2.8	847	1.79	75.1
O	Wat-131	O	Glu-19	3.1	336	1.96	43.9

to water, which is actually somewhat lower than the 3–4 solute–solvent hydrogen bonds found previously in a simulation of formamide in water.[43]

(C) **Comparison with Water in the Crystal.** The water molecules that show particularly low mobility in Figure 9 are near the protein. In order to understand better the reasons for their atypical behavior, a preliminary selection of those water molecules in the lower range of RMS fluctuations was made. Analysis of the hydrogen bonding for these water molecules during the final 70 ps revealed that most of them have one or more hydrogen bonds with the protein that are formed a high percentage of the time. Not surprisingly, many of these water molecules were located in the crystal structure,[22] though again the crystallographic water was not used in the simulation setup. Table IV contains a list of the water molecules from the simulation that have RMS fluctuations of 3.8 Å or less and form hydrogen bonds to the same protein atoms as observed for the crystalline water for at least 40% of the simulation. By these criteria, out of a total of 30 water molecules that form hydrogen bonds to the protein in the crystal, 19 are reproduced in the molecular dynamics calculation. Small shells around some of these molecules from an arbitrarily selected, instantaneous structure at 80 ps are compared to the corresponding segments of the crystal structure in Figure 13. Many of the low-mobility water molecules are found to form bridges between hydrogen-bonding units in different residues, essentially filling in the protein structure.

Conclusions

The results described above have established that an equilibrium state is attainable in a computationally accessible period of time for a small protein in water, starting from a crystal structure. Furthermore, interesting differences between the crystal and solution structures for the third domain of silver pheasant ovomucoid have been observed, most notably the extension of side chains into the solvent, the displacement of Ala-44 from the α-helix and Lys-29, Thr-30, Tyr-31, and Gly-32 out of hydrogen-bonding distance in the β-sheet, and the shifting of the Glu-19 carboxylate interaction from Thr-17 to the side chain of Arg-21. Some of these differences can be correlated to the change of environment between the crystal and aqueous solution. The calculated solution structure, the solvent dynamics, and the hydration of the protein have also been compared with available experimental and theoretical data. Further analysis of this type will contribute to the greater understanding of the interactions of proteins with the aqueous environment and the effects of these interactions on both protein structure and reactivity.

Acknowledgment. Gratitude is expressed to the National Institutes of Health for support and to Prof. Michael Laskowski, Jr., for many valuable discussions and for making available the data in ref 24 prior to publication.

Registry No. Water, 7732-18-5.

(43) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 1489.